STIC Search Report
Biotech-Chem Library

USPTO

STIC Database Tracking Number: 131993

TO: Celine Qian
Location: REM-2Ad4&2C70
Art Unit: 1636
Monday, September 13, 2004

Case Serial Number: 09/939209

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

sira
Search and Information
Resources Administration

This Page Blank (uspto)

If more than one search is submitted, please prioritize searches in order of need.

* *

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _Methods & systems for facilitating the diagnosis & Treatment of Schizophrenia_

Inventors (please provide full names): _Levitt et al_

Earliest Priority Filing Date: _8/24/01_

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number

Please search 4110 bp - 12150 bp of SEQ ID NO.3.
both prior art & interference.

3 - A9 - 20300
mej

- -

This Page Blank (uspto)

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

This Page Blank (uspto)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: September 11, 2004, 08:12:38 ; Search time 1835 Seconds
                    (without alignments)
                    18615.673 Million cell updates/sec

Title: US-09-939-209A-3_COPY_4110_12150
Perfect score: 8041
Sequence: 1 caaatcgaagtcctggcac.........ctaggatctcaagtggatcc 8041

Scoring table: IDENTITY NUC
                    Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing :   Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Database :   N_Geneseq_29Jan04:*
                    1: geneseqn1980s:*
                    2: geneseqn1990s:*
                    3: geneseqn2000s:*
                    4: geneseqn2001as:*
                    5: geneseqn2001bs:*
                    6: geneseqn2002s:*
                    7: geneseqn2003as:*
                    8: geneseqn2003bs:*
                    9: geneseqn2003cs:*
                    10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8041	100.0	20300	6	ABK47337	Abk47337 Genomic n
2	277.6	3.5	1238	4	AAL26766	Aal26766 Human bre
3	274.4	3.4	3165	6	ABX63113	Abx63113 Human neu
4	274.4	3.4	3167	9	ADD12660	Add12660 Human CDN
5	180.2	2.2	1902	4	AAK94664	Aak94664 Human ful
6	170.2	2.1	110000	6	ABT00010_10	Abt00010_10
7	170.2	2.1	110000	6	ABT01503_10	Abt01503_10
8	169.6	2.1	110000	4	AAK95240_10	Aak95240_10
9	169.6	2.1	110000	4	AAK96733_10	Aak96733_10
10	168.4	2.1	7457	4	AAK86211	Aak86211 Human imm
11	167.8	2.1	401	4	AAK96142	Aak96142 Human neu
12	167.8	2.1	401	4	AAK97635	Aak97635 Human neu
13	167.8	2.1	401	4	ABC00912	Abc00912 Human neu
14	167.8	2.1	401	6	ABT02405	Abt02405 Human neu
15	166.2	2.1	847	4	AAK91687	Aak91687 Human cDN
16	166.2	2.1	847	4	AAK93880	Aak93880 Human cDN
17	163.2	2.0	34337	4	ABL55857	Abl55857 Human GAB
18	162.2	2.0	414	8	ACH30357	Ach30357 Human tes
19	158.2	2.0	110000	6	ABS55320_2	Abs55320 Genomic D
20	158.2	2.0	341511	6	ABS55200	Abs55200 Genomic D
21	155.2	1.9	84607	2	AAX90847	Aax90847 Human PAC
22	154.8	1.9	16310	3	AAA34964	Aaa34964 Human ade
23	154.8	1.9	16310	3	AAF21086	Aaf21086 Human low
24	154.8	1.9	16310	7	AB296780	Ab296780 Human nuc
25	154.8	1.9	17634	3	AAA34965	Aaa34965 Human ade
26	154.8	1.9	17634	3	AAF21087	Aaf21087 Human low
27	154.8	1.9	17634	7	AB296781	Ab296781 Human nuc
28	154.8	1.9	81800	7	ABK84756	Abk84756 Human CDN
29	153.8	1.9	110000	6	ABL57909_0	Abl57909 Human tra
30	150.2	1.9	222880	6	ADC87622	Adc87622 Human GPC
31	147.6	1.8	126512	6	ABN83429	Abn83429 Human tra
32	147.2	1.8	110000	6	ABS55320_0	Abs55320 Human tra
33	147.2	1.8	341511	6	ABS55200	Abs55200 Genomic D
34	146.8	1.8	110000	6	ABQ74964_1	Abq74964 Continuation (2 of
35	146	1.8	334462	6	ADC24763	Adc24763 Human wil
36	144.8	1.8	326014	6	ABK89296	Abk89296 Human gen
37	144.4	1.8	140167	6	ABT10146	Abt10146 Human bre
38	142.8	1.8	110000	4	AAK95240_03	Aak95240_03 Continuation (4 of
39	142.8	1.8	110000	6	AAK96733_03	Aak96733_03 Continuation (4 of
40	142.8	1.8	110000	6	ABT00010_03	Abt00010_03 Continuation (4 of
41	142.8	1.8	110000	6	ABT01503_03	Abt01503_03 Continuation (4 of
42	142	1.8	7163	4	AAS46545	Aas46545 Tumour su
43	141.2	1.8	349881	9	ADC86642	Adc86642 Human GPC
44	141	1.8	41684	3	AAA28150	Aaa28150 Human pur
45	140.4	1.7	175590	9	ADD50650	Add50650 BAC seque

ALIGNMENTS

RESULT 1
ID   ABK47337 standard; DNA; 20300 BP.
XX
AC   ABK47337;
XX
DT   18-JUN-2002 (first entry)
XX
DE   Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW   RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW   neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW   gene; ds.
XX
OS   Homo sapiens.
XX

FH   Key   Location/Qualifiers
FH
FT   variation   replace(4121,T)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= a
FT   variation   replace(4123,A)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= b
FT   variation   replace(4368,C)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= c
FT   variation   replace(4621,C)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= d
FT   variation   replace(4790,T)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= e
FT   variation   replace(4816,T)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= f
FT   variation   replace(4970,T)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= g
FT   variation   replace(5055,G)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= h
FT   variation   replace(5295,A)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= i
FT   variation   replace(5695,A)
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /tag= j

FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(7375,T)
FT   /*tag= k
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(7759,A)
FT   /*tag= l
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(8559,A)
FT   /*tag= m
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(8596,A)
FT   /*tag= m
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(9602..9610,CA)
FT   /note= "Deletion polymorphism"
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(9892,A)
FT   /*tag= n
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(9963,A)
FT   /*tag= o
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(10132,A)
FT   /*tag= p
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(10156,C)
FT   /*tag= q
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(11091,T)
FT   /*tag= r
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(11106,A)
FT   /*tag= s
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(11774,T)
FT   /*tag= t
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(12143,A)
FT   /*tag= u
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(12145,T)
FT   /*tag= v
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(14367,G)
FT   /*tag= w
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(17027..17029,GA)
FT   /*tag= x
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /replace(17630,T)
FT   /note= "Deletion polymorphism"
FT   /*tag= y
FT   variation
FT   /standard_name= "Single-nucleotide polymorphism"
FT   /*tag= z

CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level. The
CC   method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This method
CC   offers the possibility of diagnosing schizophrenia by a biological test
CC   and no longer exclusively by clinical evaluations. The present sequence
CC   represents the genomic DNA encoding the human regulator of G-protein
CC   signalling 4 (RGS4) protein used in the method of the invention. The gene
CC   for the RGS4 protein is located on human chromosome 1q21-22
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 U; 0 Other;
XX

Query Match   100.0%; Score 8041; DB 6; Length 20300;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 8041; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 CAAATTGAAGTCCTTGGGCACGTTACAAAACTAGGAGGTTCAATAAAAGTTTACCCTAATTTACC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4110 CAAATTGAAGTCCTTGGGCACGTTACAAAACTAGGAGGTTCAATAAAAGTTTACCCTAATTTACC 4169

Qy   61 AAGATTTCCTAGAGGATCTATAATTGGAATCCAGATCTGCCCCTTCTCTGTAAAGTTCAAGCA 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4170 AAGATTTCCTAGAGGATCTATAATTGGAATCCAGATCTGCCCCTTCTCTGTAAAGTTCAAGCA 4229

Qy   121 CTTTCCATGACACCATACTGTTTCTTTCCTTCCACCTGCACAAATGCAAATGAACTCTTATGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4230 CTTTCCATGACACCATACTGTTTCTTTCCTTCCACCTGCACAAATGCAAATGAACTCTTATGAAA 4289

Qy   181 CTGCTGTTTCTATCCTGGGCTAAATGTTCAGAAAAAAGAGATTTAAACTTTGGGATAAGGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4290 CTGCTGTTTCTATCCTGGGCTAAATGTTCAGAAAAAAGAGATTTAAACTTTGGGATAAGGC 4349

Qy   241 TATTTTGGGTTTTCTTCCTTCTTCTTCTTCGGGAAAACAAAGGTTTTTCTTTCCCCTGGTAATTAAGT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4350 TATTTTGGGTTTTCTTCCTTCTTCTTCTTCGGGAAAACAAAGGTTTTTCTTTCCCCTGGTAATTAAGT 4409

Qy   301 GTGGTATTGTTCTTCCAGGGAAATCAGTGATGCATCACCTGCTGCTATCAAATGTTCAGGG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4410 GTGGTATTGTTCTTCCAGGGAAATCAGTGATGCATCACCTGCTGCTATCAAATGTTCAGGG 4469

Qy   361 TTGGAGTTCCTGATTTATTGCATGTGGCCCACAAAGCTTGGTCGTGCAAAGAATTGGACACATT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4470 TTGGAGTTCCTGATTTATTGCATGTGGCCCACAAAGCTTGGTCGTGCAAAGAATTGGACACATT 4529

Qy   421 TCCCAAAAGTAAGACATACTGGGAAGTCCCTGTTTACCTTCCCTGGTATACAGCATCCTCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4530 TCCCAAAAGTAAGACATACTGGGAAGTCCCTGTTTACCTTCCCTGGTATACAGCATCCTCC 4589

Qy   481 AGCCCCCATATCTTTGCTTTTTAGTCCTAAAAATCAATAACTCTCATTCATTGATTGTTCTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4590 AGCCCCCATATCTTTGCTTTTTAGTCCTAAAAATCAATAACTCTCATTCATTGATTGTTCTA 4649

Qy   541 GGCCATTGTAGTAAACAATAAAGGAAGGAGGGAGGGCTTCTGACAAACTGAGAGAAAATTGTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4650 GGCCATTGTAGTAAACAATAAAGGAAGGAGGGAGGGCTTCTGACAAACTGAGAGAAAATTGTC 4709

Qy   601 ATCTGAAGTGGTGCAAGCACAGCCTGGGGCTGAGCCTTGGCCTACATCCTGCCCCAAGTGG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4710 ATCTGAAGTGGTGCAAGCACAGCCTGGGGCTGAGCCTTGGCCTACATCCTGCCCCAAGTGG 4769

Qy   661 AGGATGCAGTGCCCATTTAACATCTGGTAGAAGTAAAGCGAACGGCTGCCACAATGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4770 AGGATGCAGTGCCCATTTAACATCTGGTAGAAGTAAAGCGAACGGCTGCCACAATGA 4829

Qy   721 CTTATTTCCTGCATTTGATACACCGTCAATCCTTGAGAAAGTGTTTTCTTTTGTTCTCCCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4830 CTTATTTCCTGCATTTGATACACCGTCAATCCTTGAGAAAGTGTTTTCTTTTGTTCTCCCTG 4889

Qy   781 AGCAAAGGTTGGAAAAATTGAAAAATTTACCTAGAGACCACACAATAGTTCATCCTCTG 840

PT   Diagnosing, assessing susceptibility and treating schizophrenia, involves
PT   observing regulator of G-protein signaling 4, RGS4 levels in a subject.
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX

CC   This invention relates to a novel method for diagnosing schizophrenia or
CC   determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,

```
Db  4890 AGCAAAAGGTTGGAAAAATTGAAAATTACCTACACCACCATAGTTCACATCCTCCTG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   ... 4949

Qy   841 TGTGGCTGAATGTTGCCCCCAGTAGGAGAAACAGTTCTTCTTCAAAGCCTATTGTCAACAAT  900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db

Db  4950 TGTGGCTGAATGTTGCCCCCAGTAGGAGAAACAGTTCTTCTTCAAAGCCTATTGTCAACAAT 5009
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy
```

(DNA sequence alignment — rows alternately labelled Qy and Db with coordinate numbers on each side; identity bars shown between aligned rows. Two-column layout continues with coordinate ranges 841–7109.)

```
Qy    7381  TGCAAATAGAAATTACCACAAATTTTTATTTTTGGAAATAATATCTTCAGAAAACATAATTTTTT  7440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11490  TGCAAATAGAAATTACCACAAATTTTTATTTTTGGAAATAATATCTTCAGAAAACATAATTTTTT  11549

Qy    7441  ATGTACTATTAAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCG  7500
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11550  AGTACTATTAAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCG  11609

Qy    7501  ATGGGCTTCTTTAAAATGAACCAGGAGGGGTCTGGGCAGAGAGAGTATCTTATGAAATAAGTTCCT  7560
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11610  ATGGGCTTCTTTAAAATGAACCAGGAGGGGTCTGGGCAGAGAGAGTATCTTATGAAATAAGTTCCT  11669

Qy    7561  CTTGACCTTCAGCGCTTAATTTTCTGAATGAGGAGTGGAGCAAATTACTTCAAGCTTCACTTAA  7620
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11670  CTGACCTTCAGCGCTTAATTTTCTGAATGAGGAGTGGAGCAAATTACTTCAAGCTTCACTTAA  11729

Qy    7621  CTTGCATATGAAATGAAACCCTACAAACGTTACAACAGAGTGTCACAGAGAAAGTTATGCTCTGG  7680
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11730  CTTGCATATGAAATGAAACCCTACAAACGTTACAACAGAGTGTCACAGAGAAAGTTATGCTCTGG  11789

Qy    7681  TAAATATTTTTGCAAAACAGATAAAAAGATAATAACTACTAGGAGCTCTCGTCCCTCAAAGAGTTAAGC  7740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11790  TAAATATTTTTGCAAAACAGATAAAAAGATAATAACTACTAGGAGCTCTCGTCCCTCAAAGAGTTAAGC  11849

Qy    7741  AGCTAAATCTAAGGAGGTAAACTCTATGTCAGCAGGAGAACTGCTCTCTTTCCCTTTCCTTCCT  7800
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11850  AGCTAAATCTAAGGAGGTAAACTCTATGTCAGCAGGAGAACTGCTCTCTTTCCCTTTCCCCT  11909

Qy    7801  CAATAAAATTGCAAAATCATCTAGTCCAAACATCTTTACCACCAGTGCCTGAGGCCTCCAGAGG  7860
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11910  CAATAAAATTGCAAAATCATCTAGTCCAAACATCTTTACCACCAGTGCCTGAGGCCTCCAGAGG  11969

Qy    7861  AGCCATTGCCTTCTCAAGGTCACATAGGTGGTGTGGGTGACTTAGGACACAAGAGAATGGAAAGTTGTGAATTC  7920
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11970  AGCCATTGCCTTCTCAAGGTCACATAGGTGGTGTGGGTGACTTAGGACACAAGAGAATGGAAAGTTGTGAATTC  12029

Qy    7921  CTGACTCCAGTAACTTCTCTGAAGTCATTTTGTTTTTTGTTTTATTTTATTTATTATAAGAA  7980
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12030  CTGACTCCAGTAACTTCTCTGAAGTCATTTTGTTTTTTGTTTTATTTTATTTATTATAAGAA  12089

Qy    7981  TACTTGCTAAGCACACTTACCCCTCCTGCCATTGATTAATAACTCTTAGGAATCTCAGGTGGATC  8040
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12090  TACTTGCTAAGCACACTTACCCCTCCTGCCATTGATTAATAACTCTTAGGAATCTCAGGTGGATC  12149

Qy    8041  C  8041
            |
Db   12150  C  12150
```

```
Query Match             3.5%;  Score 277.6;  DB 4;  Length 1238;
Best Local Similarity   98.6%;  Pred. No. 2e-54;
Matches   280;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    6205  CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAAGGAGCAGCTTGGTACTTCAGGACGCGGTGTCT  6264
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      58  CTCCTTAAGGGCTGGAGAGGCAGAGGGAGACAAGGAGCAGCTTGGTACTTCAGGACGCGGTGTCT  117

Qy    6265  GATTGGCTGGACGGTCGTAGCTGGGCTGGGCTATAAAAGAGACAGCCCTTACAGCCTTAGCAGGAAGA  6324
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118  GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACAGCCCTTACAGCCTTAGCAGGAAGA  177

Qy    6325  CGGCTCAGAGGATTCGTCACAAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA  6384
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178  CGGCTCAGAGGATTCGTCACAAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA  237

Qy    6385  AGCCACAGCTCTCCCCTGCCGCATTTCTTTCCCGCTTCGGAATTCCAAGCTGTTAAATAAG  6444
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     238  AGCCACAGCTCTCCCCTGCCGCATTTCTTTCCCGCTTCGGAATTCCAAGCTGTTAAATAAG  297

Qy    6445  ATGTGCAAAGGGCTTGCAGGTCTGCCGGGCTTCCTTGCTTGAGGAG  6488
            ||||||||||||||||||||||||||||||||||||||||||||||
Db     298  ATGTGCAAAGGGCTTGCAGGTCTGCCGGGCTTCCTTGCTTGAGGAG  341
```

```
XX
PF   08-JAN-2002; 2002US-00044090.
XX
PR   28-JUL-2000; 2000US-0224069P.
PR   08-JAN-2001; 2001US-0260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PS   Claim 1; Page; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that are
CC   differentially expressed in activated vascular tissue. The invention also
CC   discloses a high throughput method for detecting differentially expressed
CC   cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a high-
CC   throughput methods for detecting differential expression of one or more
CC   cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify an antibody to the protein that can
CC   be used to detect a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   diagnosing, staging, treating, or monitoring the progression of treatment
CC   of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC   disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
CC   injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC   genetic or gene expression analysis of several new nucleic acid
CC   molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC   diagnosing pre-pathologic disorders, and chronic or acute diseases
CC   associated with abnormalities in the expression, amount or distribution
CC   of the protein. The present sequence represents a cDNA of the invention
CC   that is differentially expressed in activated vascular tissue. Note: The
CC   sequence data for this patent did not form part of the specification, but
CC   was obtained in electronic format directly from USPTO at
CC   http.segdata.uspto.gov/sequence.html?DocID=2002013708I
XX
SQ   Sequence 3165 BP; 889 A; 636 C; 672 G; 967 T; 0 U; 1 Other;

Query Match    3.4%;  Score 274.4;  DB 7;  Length 3165;
Best Local Similarity    97.9%;  Pred. No. 1.7e-53;
Matches  278;  Conservative  0;  Mismatches  6;  Indels  0;  Gaps  0;

Qy      6205 CTTTCAGGGGCTGGAGGAGGCAGAGGGAGACAGAGGAGAGCTGGTACTGCAGAGCGGTCGCT 6264
             ||||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63 CGTAGAGGGGCTCGGAGGAGGCAGAGGGAGAGCTGGTACTGCAGAGCGGTCGTCT 122

Qy      6265 GATTGGCTGACACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACCAGGGCTTAGCAGGAAGA 6324
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       123 GATTGGCTGGACGGGTCGTAGCTGGGCTATAAAAGAGACCCCTACCAGGGCTTAGCAGGAAGA 182

Qy      6325 CGCTCAGAGGATTCTGACAATATCTTTACGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 6384
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       183 CGCTCAGAGGATTCTGACAATATCTTTACGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 242

Qy      6385 AGCCACAGCTCCTCCTGCGCCGCATTTCTTCCTGCTTGGAATTCCAAGCTGTTAAATAAG 6444
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       243 AGCCACAGCTCCTCCTGCGCCGCATTTCTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 302

Qy      6445 ATGTGCAAAAGGGCTTGCAGGTCTGCCGGGGCTTCAGGGTGTCTGCCTTGAGGAG 6488
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       303 ATGTGCAAAAGGGCTTGCAGGTCTGCCGGGGCTTCAGGGTGTCTGCCTTGAGGAG 346
```

```
RESULT 4
AD   AD12660
ID   AD12660 standard; cDNA; 3167 BP.
XX
AC   AD12660;
XX
DT   01-JAN-2004  (first entry)
XX
PN   US2003113720-A1.
XX
PD   19-JUN-2003.
XX
PF   30-JUL-2001; 2001US-00918624.
XX
PR   28-JUL-2000; 2000US-0224070P.
XX
PA   (SCHE/) SCHEBYE X M.
PA   (SORN/) SORNASSE T.
XX
PI   Schebye XM, Sornasse T;
XX
DR   WPI; 2003-810888/76.
XX
PN   US2003113720-A1.
XX
DE   Human cDNA expressed during adipocyte differentiation #6.
XX
KW   ssi; gene; human; adipocyte differentiation; obesity; type II diabetes;
KW   lipodystrophy; hyperinsulinaemia.
XX
OS   Homo sapiens.
XX
FH
XX
PS   Claim 1; SEQ ID NO 6; 105pp; English.
XX
CC   Novel isolated cDNAs expressed in adipocyte differentiation useful for
CC   treating subject with disorder such as obesity, type II diabetes,
CC   lipodystrophy or hyperinsulinemia. The invention relates to an isolated cDNA expressed during adipocyte
CC   differentiation. The cDNA is useful for treating a subject with a
CC   disorder such as obesity, type II diabetes, lipodystrophy or
CC   hyperinsulinaemia. The nucleic acid is useful for a high throughput
CC   method of using a cDNA to screen several molecules or compounds to
CC   identify a ligand which specifically binds the cDNA which involves
CC   combining the nucleic acid with several molecules or compounds under
CC   conditions to allow specific binding, and detecting specific binding
CC   between each cDNA and at least one molecule or compound, thus identifying
CC   a ligand that specifically binds to each cDNA. The several molecules or
CC   compounds are chosen from DNA molecules, RNA molecules, peptide nucleic
CC   acid molecules, mimetics, peptides, transcription factors, repressors and
CC   regulatory proteins. The present sequence represents a human cDNA
CC   expressed during adipocyte differentiation.
XX
SQ   Sequence 3167 BP; 877 A; 625 C; 671 G; 970 T; 0 U; 24 Other;

Query Match    3.4%;  Score 274.4;  DB 9;  Length 3167;
Best Local Similarity    97.9%;  Pred. No. 1.7e-53;
Matches  278;  Conservative  0;  Mismatches  6;  Indels  0;  Gaps  0;

Qy      6205 CTTTCAGGGGCTGGAGGAGGCAGAGGGAGACAGAGGAGAGCTGGTACTGCAGAGCGGTCGCT 6264
             ||||       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63 CGTAGAGGGGCTCGGAGGAGGCAGAGGGAGAGCTGGTACTGCAGAGCGGTCGTCT 122

Qy      6265 GATTGGCTGACACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACCAGGGCTTAGCAGGAAGA 6324
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       123 GATTGGCTGGACGGGTCGTAGCTGGGCTATAAAAGAGACCCCTACCAGGGCTTAGCAGGAAGA 182

Qy      6325 CGCTCAGAGGATTCTGACAATATCTTTACGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 6384
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       183 CGCTCAGAGGATTCTGACAATATCTTTACGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 242

Qy      6385 AGCCACAGCTCCTCCTGCGCCGCATTTCTTCCTGCTTGGAATTCCAAGCTGTTAAATAAG 6444
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   243 AGCCACAGCTCCTCCTGCGCCGCATTCTTTCCTGCTTGCGCAATTCCAAGCTGTTAAATAAG 302

Qy  6445 ATGTGCAAAGGCGCTTGCCAGCTCTGCCGGCTTCTTGCTTGAGGAG 6488
Db   303 ATGTGCAAAGGCGCTTGCCAGCTCTGCCGGCTTCTTGCTTGAGGAG 346

RESULT 5
AAK94664
ID   AAK94664 standard; cDNA; 1902 BP.
XX
AC   AAK94664;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human full-length cDNA, SEQ ID NO: 3657.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-00114089.
XX
PR   08-JUL-1999;  99JP-00194486.
PR   11-JAN-2000;  2000JP-00118774.
PR   02-MAY-2000;  2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Nishikawa T, Isogai T, Hayashi K, Kawai Y;
PI   Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR   WPI; 2001-524255/58.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their use
PT   in genetic manipulation.
XX
PS   Claim 8; SEQ ID NO 3657; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA. The present sequence is a
CC   methods. The present sequence is a full length human cDNA of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in CD-ROM format directly
CC   from EPO
XX
SQ   Sequence 1902 BP; 576 A; 333 C; 440 G; 553 T; 0 U; 0 Other;

Query Match          2.2%;  Score 180.2;  DB 4;  Length 1902;
Best Local Similarity 66.8%;  Pred. No. 1.5e-31;
Matches   404;  Conservative   0;  Mismatches  168;  Indels   33;  Gaps    9;

Qy  1245 CAGGTTAATAAGCAAAACCTACACAAAATTTATTTAATCAAAGTTTTTTACATGACATGGA 1304
Db   213 CAAGTTAATAACAGAAAAGGTATAACAAA----TTAATCAAAGATTTATGTGACCCAAGA 267

Qy  1305 GTCTCTCAGAAAATGAGAGAACCAGAAGACCCAGGGAAAGTTGTCTGTTTTTTTGCTGAGGTT 1364
Db   268 GCCTTCAAAAATGAGAGGACCCAAAGAGTGTAGGGAAAAGCTACCTACCTCGTTTTTCT--GCTAAGTT 325

Qy  1365 CGATGAAGAATGGATAGAGCATGTAGCCATGTAGGATTAGACAAAGGATAAGGATAAGATGATCTAGTGGT 1424
Db   326 TAACGTAGAATGGACAGCCGTCTAGAAAATGCGACTGGACAAAAGGAATATGATTGCTTGAC 385
```

```
Qy  1425 AAAGGACTCAGGGGGGAAAACACAGAAGAGGCCTGCTGTCTATTCGACAAGGCTGTTGGGACAGGACTGGAATGGAGGGT 1484
Db  1425 ...TGTGTAACATTTCTTCCTCCCCGTTATAGGGCACAGCCCCACTGGAATGGAGGGT 1484

Qy    443 ------TGTGTAACATTTCTTCCTCCCCGTTATAGGGCACAGCCCCACTGGAATGGAGGGT 496

Qy  1485 TCTCTATGTATAGACATTCTTTCCTCCTGAGTATGGGGGCAGGAGACTTCTTCAATGAGGGT 1544
Db    386 AATAGACTGAGAGTGTGGCAAAACCGA--GCCAGCGTGTCCAGACACTCTTCTTTGGTCTCTC-- 442

Qy  1545 CTTCAAGGGAGGAAGGGAGAGGGAGAAGTGG----------CCTTTTTAGATTTATGGGTGGCTTCG 1596
Db    497 CTTCAAGGGAGGAAGGGAGAGGGAGAAGTAAC CTTCAGAGAAGTTTTATGGACTTGGTGTG 556

Qy  1597 GGGA-AGAGGAGTTCTAGTTTCTATGACCCATCTGGGCGAGAAGGAAGGGAATTCTGGTTTCGT 1655
Db    557 GGAAGAGGGGGAATTCAAGTGTCTATGACCTGCCTTGGGACGAAGACGGAAATTTTATTCCTTAT 616

Qy  1656 GA-CTTGCTTTCAGAAGAAAGAGGAGTAAGAGGCAGGAGGCGCAGAAGGACGAGGTCAGAAAG 1714
Db    617 GAACTCACTATGGGGGCAGAAGGGAGAAGGGGCTGGAGAGAAGGGCAGCAG 676

Qy  1715 AGACTTGGCGTGCTTCTGAGGGCTTCCGGCTCTCCTTTAGTTCCAAGTACTTTCTTAGCATAC 1774
Db    677 AGGCTTTGC----TTTGAGGGCTTCTTCAATTTCCTTTAGTTCAAAG---TCCTTGGGATGC 729

Qy  1775 CAAAGCACTATACTTTGGCAATAGGTTTTCTGAGCTCTAACACTGCAATCATGCTAAACT 1834
Db    730 CAAAGCATTATATTTTGGAGTAATTCTCTTCTAAGGCACAACAAACTAAAAAAATGAAAAT 789

Qy  1835 CCTCT 1839
Db    790 CATTT 794

RESULT 6
ABT00010_10
Continuation (11 of 15) of ABT00010 from base 1000001 (Human neuregulin 1 gene.)
WP Sequence split into 15 fragments
WP Fragment Name          Begin          End            LOCUS ABT00010 Accession Abt00010
WP ABT00010_00            1              110000
WP ABT00010_01            100001         210000
WP ABT00010_02            200001         310000
WP ABT00010_03            300001         410000
WP ABT00010_04            400001         510000
WP ABT00010_05            500001         610000
WP ABT00010_06            600001         710000
WP ABT00010_07            700001         810000
WP ABT00010_08            800001         910000
WP ABT00010_09            900001         1010000
WP ABT00010_10            1000001        1110000
WP ABT00010_11            1100001        1210000
WP ABT00010_12            1200001        1310000
WP ABT00010_13            1300001        1410000
WP ABT00010_14            1400001        1503841

Query Match          2.1%;  Score 170.2;  DB 6;  Length 110000;
Best Local Similarity 68.8%;  Pred. No. 1.9e-28;
Matches   232;  Conservative   1;  Mismatches  104;  Indels   0;  Gaps    0;

Qy  3006 TTTCTCACATCCCGTGACTTCTGTGCCCCTGGGTATCCATCATCACCTGAATGATGTGCATT 3065
Db 13790 TTTGTTACATGGATATAACTCGGTCTTTCAGTGTAACCATCACTCAAAATAGTGTACCTT 13849

Qy  3066 GTACCCTTAAGTAAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCCTCGTCGAGGC 3125
Db 13850 GTACCTATTACGTAACTTGTCATTCCTCACCCCTCCACCCTCCATCTTTYCTAGTC 13909

Qy  3126 TCCATTGTCCATCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCACTTA 3185
Db 13910 TCCAATGTCTATATTATTTCCACCCTCTA=GTCCATGTGTACTCATTATTTAGGTCTACTTA 13969

Qy  3186 TAAGTGATAACATGCAAATATTTGTCTTTCTGTGTCTGTTGTTTACTTATGGATTAATGG 3245
Db 13969 ... 
```

RESULT 7

ABT01503_10

Continuation (11 of 15) of ABT01503 from base 1000001 (Human neuregulin 1 gene.)

WP Sequence split into 15 fragments   LOCUS ABT01503 Accession Abt01503

WP Fragment Name	Begin	End
WP ABT01503_00	1	110000
WP ABT01503_01	100001	210000
WP ABT01503_02	200001	310000
WP ABT01503_03	300001	410000
WP ABT01503_04	400001	510000
WP ABT01503_05	500001	610000
WP ABT01503_06	600001	710000
WP ABT01503_07	700001	810000
WP ABT01503_08	800001	910000
WP ABT01503_09	900001	1010000
WP ABT01503_10	1000001	1110000
WP ABT01503_11	1100001	1210000
WP ABT01503_12	1200001	1310000
WP ABT01503_13	1300001	1410000
WP ABT01503_14	1400001	1503841

Query Match   2.1%; Score 170.2; DB 6; Length 110000;
Best Local Similarity 68.8%; Pred. No. 1.9e-28;
Matches 232; Conservative 1; Mismatches 104; Indels 0; Gaps 0;

Qy 3006 TTTCTCACCATCCGCTGACTTCTTTGCCCCTGGGGTATTCATCACCTGAATGATGTGCATT 3065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13790 TTTGTTACATGGATAATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 13849

Qy 3066 GTACCCCTTAAGTAAGTTAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTAGGC 3125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13850 GTACCTATTACGTTAACTTGTCATTCCTCACCCTCTCCACCCTCCAATCTTTNTGAGTC 13909

Qy 3126 TCCAATTGTCCATCAATTCCACACTCTACACATCTATGTGTACACATTATTTAGGCTCCTACTTA 3185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13910 TCCAATGTCTATTATTCCACCCTCTTATGTCCTGTTGTTCTGTTTTACTTAAGATAATAG 13969

Qy 3186 TAAGTGATAAGCATGCAAATAATTGTCTTCTGTGTCTGTTGTTCTGTTTTACTTAAGATAATGG 3245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13970 TAAGTAAGAACATGCAAATAATTGACTTGCTGTTATTTTCTGTTTTCACTTAAGCTGAATG 14029

Qy 3246 CCCCCAGTTCTATCTCAGGCTGCTGCAAAAGGCAGATGATTTCATTGTTTTTTTATGGCTATGT 3305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 14030 CCTTCAGGTCCATCCAATGTTCTGCAAAAGACAATGAATTGATTCTTTTGTATGGCTAATG 14089

Qy 3306 TCTTTCCCAAATTTAGATAAAGAACACTCGCACTTGCT 3342
          ||||||||||||||||||||||||||||||||||||
Db 14090 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 14126

RESULT 8

AAK95240_10

Continuation (11 of 15) of AAK95240 from base 1000001 (Human neuregulin-1 gene.)

WP Sequence split into 15 fragments   LOCUS AAK95240 Accession Aak95240

WP Fragment Name	Begin	End
WP AAK95240_00	1	110000
WP AAK95240_01	100001	210000
WP AAK95240_02	200001	310000
WP AAK95240_03	300001	410000
WP AAK95240_04	400001	510000
WP AAK95240_05	500001	610000
WP AAK95240_06	600001	710000
WP AAK95240_07	700001	810000
WP AAK95240_08	800001	910000
WP AAK95240_09	900001	1010000
WP AAK95240_10	1000001	1110000
WP AAK95240_11	1100001	1210000
WP AAK95240_12	1200001	1310000
WP AAK95240_13	1300001	1410000
WP AAK95240_14	1400001	1503900

Query Match   2.1%; Score 169.6; DB 4; Length 110000;
Best Local Similarity 68.8%; Pred. No. 2.6e-28;
Matches 232; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 3006 TTTCTCACCATCCGCTGACTTCTTTGCCCCTGGGGTATTCATCACCTGAATGATGTGCATT 3065
Db 13849 TTTGTTACATGGATAATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 13908

Qy 3066 GTACCCCTTAAGTAAGTTAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTAGGC 3125
Db 13909 GTACCTATTACGTTAACTTGTCATTCCTCACCCTCTCCACCCTCCAATCTTTNTGAGTC 13968

Qy 3126 TCCAATTGTCCATCAATTCCACACTCTACACATCTATGTGTACACATTATTTAGGCTCCTACTTA 3185
Db 13969 TCCAATGTCTATTATTCCACCCTCTTATGTCCTGTTGTTCTGTTTTACTTAAGATAATAG 14028

Qy 3186 TAAGTGATAAGCATGCAAATAATTGTCTTCTGTGTCTGTTGTTCTGTTTTACTTAAGATAATGG 3245
Db 14029 TAAGTAAGAACATGCAAATAATTGACTTGCTGTTATTTTCTGTTTTCACTTAAGCTGAATG 14088

Qy 3246 CCCCCAGTTCTATCTCAGGCTGCTGCAAAAGGCAGATGATTTCATTGTTTTTTTATGGCTATGT 3305
Db 14089 CCTTCAGGTCCATCCAATGTTCTGCAAAAGACAATGAATTGATTCTTTTGTATGGCTAATG 14148

Qy 3306 TCTTTCCCAAATTTAGATAAAGAACACTCGCACTTGCT 3342
Db 14149 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 14185

RESULT 9

AAK96733_10

Continuation (11 of 15) of AAK96733 from base 1000001 (Human neuregulin-1 gene.)

WP Sequence split into 15 fragments   LOCUS AAK96733 Accession Aak96733

WP Fragment Name	Begin	End
WP AAK96733_00	1	110000
WP AAK96733_01	100001	210000
WP AAK96733_02	200001	310000
WP AAK96733_03	300001	410000
WP AAK96733_04	400001	510000
WP AAK96733_05	500001	610000
WP AAK96733_06	600001	710000
WP AAK96733_07	700001	810000
WP AAK96733_08	800001	910000
WP AAK96733_09	900001	1010000
WP AAK96733_10	1000001	1110000
WP AAK96733_11	1100001	1210000
WP AAK96733_12	1200001	1310000
WP AAK96733_13	1300001	1410000
WP AAK96733_14	1400001	1503900

Query Match   2.1%; Score 169.6; DB 4; Length 110000;
Best Local Similarity 68.8%; Pred. No. 2.6e-28;
Matches 232; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 3006 TTTCTCACCATCCGCTGACTTCTTTGCCCCTGGGGTATTCATCACCTGAATGATGTGCATT 3065
Db 13849 TTTGTTACATGGATAATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 13908

Qy 3066 GTACCCCTTAAGTAAGTTAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTAGGC 3125
Db 13909 GTACCTATTACGTTAACTTGTCATTCCTCACCCTCTCCACCCTCCAATCTTTNTGAGTC 13968

Qy 3126 TCCAATTGTCCATCAATTCCACACTCTACACATCTATGTGTACACATTATTTAGGCTCCTACTTA 3185
Db 13969 TCCAATGTCTATTATTCCACCCTCTTATGTCCTGTTGTTCTGTTTTACTTAAGATAATAG 14028

Qy 3186 TAAGTGATAAGCATGCAAATAATTGTCTTCTGTGTCTGTTGTTCTGTTTTACTTAAGATAATGG 3245
Db 14029 TAAGTAAGAACATGCAAATAATTGACTTGCTGTTATTTTCTGTTTTCACTTAAGCTGAATG 14088

Qy 3246 CCCCCAGTTCTATCTCAGGCTGCTGCAAAAGGCAGATGATTTCATTGTTTTTTTATGGCTATGT 3305
Db 14089 CCTTCAGGTCCATCCAATGTTCTGCAAAAGACAATGAATTGATTCTTTTGTATGGCTAATG 14148

Qy 3306 TCTTTCCCAAATTTAGATAAAGAACACTCGCACTTGCT 3342
Db 14149 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 14185

```
Db  13969 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTCTACCTCATTATTAGTTCTTACTTA 14028
Qy   3186 TAAGTGATAAACATGCAAATATTTGTCTTCTGTGTCGTCGTTGTTTTACTTATGATAATGG 3245
           |||| ||  ||| || ||| || | ||||| | | ||| ||||| ||||  |||  ||| |
Db  14029 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGAATAATAG 14088
Qy   3246 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCCATAGATTTCATTTCTTTTTTATGCCTATGT 3305
           ||  ||| ||| |||| |||||||| ||||| |||| | ||| |||||||  ||| || || |
Db  14089 CCTTCAGGTCCATCCATGTTGCTGCGAAAAGACATGAAATTGATTCTTTTGTATGGCTGAGT 14148
Qy   3306 TCTTTCCCAATTTAGATAAAAGAAACACTCGCACTTGCT 3342
           ||||| || | | || || |||| ||| |  |  ||| |
Db  14149 AGTATTCCATGGTGTATATACGTTACATTTCTTGAT 14185

RESULT 10
AAK86211

ID   AAK86211 standard; DNA; 7457 BP.
XX
AC   AAK86211;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41023.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179905P.     PR   01-SEP-2000; 2000US-0229343P.
PR   04-FEB-2000; 2000US-0180628P.     PR   01-SEP-2000; 2000US-0229344P.
PR   24-FEB-2000; 2000US-0184664P.     PR   01-SEP-2000; 2000US-0229345P.
PR   02-MAR-2000; 2000US-0186350P.     PR   05-SEP-2000; 2000US-0229509P.
PR   16-MAR-2000; 2000US-0189874P.     PR   05-SEP-2000; 2000US-0229513P.
PR   17-MAR-2000; 2000US-0190076P.     PR   06-SEP-2000; 2000US-0230437P.
PR   18-APR-2000; 2000US-0198123P.     PR   06-SEP-2000; 2000US-0230438P.
PR   19-MAY-2000; 2000US-0205515P.     PR   08-SEP-2000; 2000US-0231242P.
PR   07-JUN-2000; 2000US-0209467P.     PR   08-SEP-2000; 2000US-0231243P.
PR   28-JUN-2000; 2000US-0214886P.     PR   08-SEP-2000; 2000US-0231244P.
PR   30-JUN-2000; 2000US-0215135P.     PR   08-SEP-2000; 2000US-0231413P.
PR   07-JUL-2000; 2000US-0216647P.     PR   08-SEP-2000; 2000US-0231414P.
PR   07-JUL-2000; 2000US-0216880P.     PR   08-SEP-2000; 2000US-0232080P.
PR   11-JUL-2000; 2000US-0217487P.     PR   08-SEP-2000; 2000US-0232081P.
PR   11-JUL-2000; 2000US-0217496P.     PR   12-SEP-2000; 2000US-0231968P.
PR   14-JUL-2000; 2000US-0218290P.     PR   14-SEP-2000; 2000US-0232397P.
PR   26-JUL-2000; 2000US-0220963P.     PR   14-SEP-2000; 2000US-0232398P.
PR   26-JUL-2000; 2000US-0220964P.     PR   14-SEP-2000; 2000US-0232399P.
PR   14-AUG-2000; 2000US-0225213P.     PR   14-SEP-2000; 2000US-0232401P.
PR   14-AUG-2000; 2000US-0225214P.     PR   14-SEP-2000; 2000US-0233063P.
PR   14-AUG-2000; 2000US-0225266P.     PR   14-SEP-2000; 2000US-0233064P.
PR   14-AUG-2000; 2000US-0225267P.     PR   14-SEP-2000; 2000US-0233065P.
PR   14-AUG-2000; 2000US-0225268P.     PR   21-SEP-2000; 2000US-0234223P.
PR   14-AUG-2000; 2000US-0225270P.     PR   21-SEP-2000; 2000US-0234274P.
PR   14-AUG-2000; 2000US-0225447P.     PR   25-SEP-2000; 2000US-0234997P.
PR   14-AUG-2000; 2000US-0225557P.     PR   25-SEP-2000; 2000US-0234998P.
PR   14-AUG-2000; 2000US-0225758P.     PR   27-SEP-2000; 2000US-0235834P.
PR   14-AUG-2000; 2000US-0225759P.     PR   27-SEP-2000; 2000US-0235836P.
PR   18-AUG-2000; 2000US-0226279P.     PR   29-SEP-2000; 2000US-0235327P.
PR   22-AUG-2000; 2000US-0226681P.     PR   29-SEP-2000; 2000US-0235367P.
PR   22-AUG-2000; 2000US-0226868P.     PR   29-SEP-2000; 2000US-0235368P.
PR   23-AUG-2000; 2000US-0227182P.     PR   29-SEP-2000; 2000US-0235369P.
PR   30-AUG-2000; 2000US-0227109P.     PR   29-SEP-2000; 2000US-0235370P.
PR   01-SEP-2000; 2000US-0229287P.     PR   02-OCT-2000; 2000US-0236802P.
                                       PR   02-OCT-2000; 2000US-0237037P.
                                       PR   02-OCT-2000; 2000US-0237038P.
                                       PR   02-OCT-2000; 2000US-0237039P.
                                       PR   02-OCT-2000; 2000US-0237040P.
                                       PR   13-OCT-2000; 2000US-0239355P.
                                       PR   20-OCT-2000; 2000US-0239937P.
                                       PR   20-OCT-2000; 2000US-0240960P.
                                       PR   20-OCT-2000; 2000US-0241221P.
                                       PR   20-OCT-2000; 2000US-0241785P.
                                       PR   20-OCT-2000; 2000US-0241786P.
                                       PR   20-OCT-2000; 2000US-0241787P.
                                       PR   20-OCT-2000; 2000US-0241808P.
                                       PR   20-OCT-2000; 2000US-0241809P.
                                       PR   20-OCT-2000; 2000US-0241826P.
                                       PR   01-NOV-2000; 2000US-0244617P.
                                       PR   08-NOV-2000; 2000US-0246477P.
                                       PR   08-NOV-2000; 2000US-0246478P.
                                       PR   08-NOV-2000; 2000US-0246455P.
                                       PR   08-NOV-2000; 2000US-0246524P.
                                       PR   08-NOV-2000; 2000US-0246525P.
                                       PR   08-NOV-2000; 2000US-0246526P.
                                       PR   08-NOV-2000; 2000US-0246527P.
                                       PR   08-NOV-2000; 2000US-0246528P.
                                       PR   08-NOV-2000; 2000US-0246532P.
                                       PR   08-NOV-2000; 2000US-0246609P.
                                       PR   08-NOV-2000; 2000US-0246610P.
                                       PR   08-NOV-2000; 2000US-0246611P.
                                       PR   17-NOV-2000; 2000US-0246613P.
                                       PR   17-NOV-2000; 2000US-0249207P.
                                       PR   17-NOV-2000; 2000US-0249208P.
                                       PR   17-NOV-2000; 2000US-0249209P.
                                       PR   17-NOV-2000; 2000US-0249210P.
                                       PR   17-NOV-2000; 2000US-0249211P.
                                       PR   17-NOV-2000; 2000US-0249212P.
```

PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249302P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0256978P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX
PS   Disclosure; SEQ ID NO 41023; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 7457 BP; 2326 A; 1283 C; 1300 G; 2548 T; 0 U; 0 Other;

Query Match   2.1%;   Score 168.4; DB 4; Length 7457;
Best Local Similarity   67.2%; Pred. No 1.6e-28;
Matches 238; Conservative   0; Mismatches 116; Indels   0; Gaps   0;

Qy   3019 GCTGACTTCTTGCCCCTGGGTATTCATCACCTGAAATGAATGTGGCATTTACCCCTTAAGT 3078
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5567 GGTGAAGTCCAGCTTCTTGGTGTATTCATCACCAAATAATATACATTGTACCCATTAAAT 5626

Qy   3079 AATTTCTTCACCATCCGGCTGACTTCTTGCCCCTCATCCTTCTGAGGCTCCATTGTCCATC 3138
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5627 AATTTCTTCATTCCCATCCCCTCCGACTCTCACCCTTCCGAGTCTTCAGTGTCTCATT 5686

Qy   3139 ATTCCACACTTCACTACTATATGTGTACACATTATTTAGCTCCTACTTATAAGTGATTAACAT 3198
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5687 ACTCCACACTTGTATGTTCATGTGTACCCATTTTTTTAGCTCTCACTGATAAAATGAGAACAT 5746

Qy   3199 GCAATATTATTTGTCTTTCTTTCGTTTCTTGTCTTGTTTTTACTTATAGATTAATGCCCCCAGTTCTAT 3258
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   5747 GCAGTATTTGACTTTCTTGTTTCTGTTGTGTTCACTTATGATAATAGACTCCAGTTCCAT 5806

Qy   3259 CTAGGCTGCTGCAAAAGGCATGAGATTCATTCATTTTTTATGGCTATGTTTTCTTTCCCCAATTT 3318
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   5807 CCATGTTGCTGCAAAAGACATGGTTTCATTCCCTCTTTTATGGCTGAATAAATATCATGAGTT 5866

Qy   3319 AGATAAAGAACACTGCCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAG 3372
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5867 CATTTAAATGCATCAATTTAATATGTATTTCAAAATGTTAAAAATGCACACTAG 5920

CC   This sequence represents a single nucleotide polymorphism (SNP) of the
CC   human neuregulin-1 associated gene 1 (NRG1AG1) of the invention. The
CC   NRG1AG1 gene is also referred to as the human Schizophrenia gene. The
CC   invention also relates to fragments or variants of the gene and the
CC   NRG1AG1 polypeptides they encode. The NRG1AG1 nucleic acids and
CC   polypeptides may be used in the prevention, diagnosis and treatment of
CC   diseases associated with inappropriate NRG1AG1 expression. For example,
CC   they may be used to treat disorders associated with decreased expression
CC   by rectifying mutations or deletions in a patient's genome that affect
CC   the activity of NRG1AG1 by expressing inactive proteins or to supplement
CC   the patients own production of NRG1AG1. Additionally, the gene may be
CC   used to produce NRG1AG1 polypeptides, by inserting the nucleic acids into
CC   a host cell and culturing the cell to express the protein. The gene may
CC   also be used as DNA probes and primers in diagnostic assays to detect and
CC   quantitate the presence of similar nucleic acids in samples, and
CC   therefore which patients may be in need of restorative therapy. The
CC   NRG1AG1 polypeptides may also be used as antigens in the production of
CC   antibodies against NRG1AG1 and in assays to identify modulators of the
CC   NRG1AG1 expression and activity. Anti-NRG1AG1 antibodies and antagonists
CC   may also be used to down regulate expression and activity. Anti-NRG1AG1
CC   antibodies may also be used as diagnostic agents for detecting the
CC   presence of NRG1AG1 polypeptides in samples. NRG1AG1 is associated with
CC   schizophrenia which may be prevented, diagnosed and/or treated by the
CC   above methods
XX

RESULT 12

```
ID   AAK97635 standard; DNA; 401 BP.
XX
AC   AAK97635;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human neuregulin 1 gene single nucleotide polymorphism SNP8NRG1013903.
XX
KW   single nucleotide polymorphism; gene therapy; SNP;
KW   Human; neuregulin 1 gene; schizophrenia;
XX
OS   Homo sapiens.
XX
XX
PN   WO20016487-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US006377.
XX
PR   28-FEB-2000; 2000US-00515716.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR   WPI; 2001-514841/56.
XX
PT   Neuregulin 1 nucleic acids and proteins useful for diagnosing preventing
PT   and treating schizophrenia.
XX
PS   Disclosure; Page 244; 756pp; English.
XX
CC   This sequence represents a single nucleotide polymorphism (SNP) from the
CC   human neuregulin 1 gene of the invention. The invention also relates to
CC   fragments or variants of the neuregulin 1 gene. The gene and its proteins
CC   may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate neuregulin 1 expression, such as
CC   schizophrenia. For example they may be used to treat disorders associated
CC   with decreased neuregulin 1 expression by rectifying mutations or
CC   deletions in a patient's genome that affect the activity of neuregulin 1
XX
SQ   Sequence 401 BP; 98 A; 79 C; 62 G; 161 T; 0 U; 1 Other;

Query Match             2.1%;  Score 167.8; DB 4; Length 401;
Best Local Similarity   71.2%;  Pred. No. 6.2e-29;
Matches  220;  Conservative    1;  Mismatches   88;  Indels    0;  Gaps    0;

Qy       3006 TTTCTCACCATCCGCTGACTTCTTGCCCCTGGGTATTCATCACCTGAATGAGTGCCATT 3065
              |||||||||||||||||||  |||||||||||||| ||| ||||||  || || ||||
Db         88 TTTGTTACATGGAATAATAACCTCTGGGTGTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy       3066 GTACCCCTTAAGTAATTCTACCATCCGGTGACTTCTTGCCCCTCATCCTTCTGAGGC 3125
              ||||| ||||  |||  |  |||||  |   | ||||   |||||||   ||| |||
Db        148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCACCCTCCCATCTTTYTGAGTC 207

Qy       3126 TCCATTGTCCATCATTCCACACTCTACATGTGTACACAATTATTTTAGCTCCTACTTA 3185
              |||| ||| |||||||||||  || | ||||||| |||||||||||||| | ||||||
Db        208 TCCAATGTCTCATTATTCCACCCTGTAGTCCAATGTGTACTCAATTATTTTAGTTCTTACTTA 267

Qy       3186 TAAGTGATAACATGCAAATATTTGTTGTTGTCTGTCTTGTTTTACTTATTATAATTGG 3245
              |||||  |||||||||||||| |||  |||||||||  | ||| |||  ||||| ||
Db        268 TAAGTAAGAACATGCAAATAATTTGACTTTCTGTTGTTGTCTGTTATTATCACTTAAGATAATAG 327

Qy       3246 CCCCCAGTTCTATTCTAGGCTGCAAAAGGCATGATTTCATTCATTTTTTTTTTATGGCTATGT 3305
              || |||| |||| |||||| |||||||| ||| || ||| || ||||||  |||||||| |||
Db        328 CCTTCAGGTCCATCCAGTGTTGCTGCAAAAAGACAATGAAATTGATTCTTTTGTATGGCTGAGT 387

Qy       3306 TCTTTCCCA 3314
              |||||  ||
Db        388 AGTATTCCA 396
```

RESULT 13

```
ID   ABT00912 standard; DNA; 401 BP.
XX
AC   ABT00912;
XX
DT   07-NOV-2002  (first entry)
XX
DE   Human neuregulin-1-associated gene 1 fragment SEQ ID NO: 941.
XX
KW   neuroleptic; gene therapy; splice variant; gene; ds.
KW   Human; neuregulin 1; neuregulin-1-associated gene 1; NRG1; NRG1AG1;
KW   schizophrenia; chromosome 8p12; single nucleotide polymorphism; SNP;
XX
OS   Homo sapiens.
XX
XX
PN   US2002045577-A1.
XX
PD   18-APR-2002.
XX
PF   28-FEB-2001; 2001US-00795668.
XX
PR   28-FEB-2000; 2000US-00515716.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
CC   by expressing inactive proteins or to supplement the patients own
CC   production of polypeptides. Additionally, the gene may be used to produce
CC   the neuregulin 1 protein, by inserting the nucleic acids into a host cell
CC   and culturing the cell to express the protein. The gene and its
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar nucleic acids in
CC   samples, and therefore which patients may be in need of restorative
CC   therapy. The protein may also be used as antigens in the production of
CC   antibodies against neuregulin 1 and in assays to identify modulators of
CC   neuregulin 1 expression and activity. The antibodies and antagonists may
CC   also be used to down regulate expression and activity. The antibodies may
CC   also be used as diagnostic agents for detecting the presence of
CC   neuregulin 1 in samples
XX
SQ   Sequence 401 BP; 98 A; 79 C; 62 G; 161 T; 0 U; 1 Other;

Query Match             2.1%;  Score 167.8; DB 4; Length 401;
Best Local Similarity   71.2%;  Pred. No. 6.2e-29;
Matches  220;  Conservative    1;  Mismatches   88;  Indels    0;  Gaps    0;

Qy       3006 TTTCTCACCATCCGCTGACTTCTTGCCCCTGGGTATTCATCACCTGAATGAGTGCCATT 3065
              |||||||||||||||||||  |||||||||||||| ||| ||||||  || || ||||
Db         88 TTTGTTACATGGAATAATAACCTCTGGGTGTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy       3066 GTACCCCTTAAGTAATTCTACCATCCGGTGACTTCTTGCCCCTCATCCTTCTGAGGC 3125
              ||||| ||||  |||  |  |||||  |   | ||||   |||||||   ||| |||
Db        148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCACCCTCCCATCTTTYTGAGTC 207

Qy       3126 TCCATTGTCCATCATTCCACACTCTACATGTGTACACAATTATTTTAGCTCCTACTTA 3185
              |||| ||| |||||||||||  || | ||||||| |||||||||||||| | ||||||
Db        208 TCCAATGTCTCATTATTCCACCCTGTAGTCCAATGTGTACTCAATTATTTTAGTTCTTACTTA 267

Qy       3186 TAAGTGATAACATGCAAATATTTGTTGTTGTCTGTCTTGTTTTACTTATTATAATTGG 3245
              |||||  |||||||||||||| |||  |||||||||  | ||| |||  ||||| ||
Db        268 TAAGTAAGAACATGCAAATAATTTGACTTTCTGTTGTTGTCTGTTATTATCACTTAAGATAATAG 327

Qy       3246 CCCCCAGTTCTATTCTAGGCTGCAAAAGGCATGATTTCATTCATTTTTTTTTTATGGCTATGT 3305
              || |||| |||| |||||| |||||||| ||| || ||| || ||||||  |||||||| |||
Db        328 CCTTCAGGTCCATCCAGTGTTGCTGCAAAAAGACAATGAAATTGATTCTTTTGTATGGCTGAGT 387

Qy       3306 TCTTTCCCA 3314
              |||||  ||
Db        388 AGTATTCCA 396
```

DR   WPI; 2002-425447/45.
XX
PT   New neuregulin 1 gene, schizophrenia gene residing in 1.5 Mb segment on
PT   human chromosome 8p12, useful for diagnosing susceptibility to or
PT   treating schizophrenia and for screening schizophrenia treating agents.
XX
PS   Disclosure; Page 600; 700pp; English.
XX
CC   The present invention relates to the human neuregulin 1 gene (NRG1),
CC   single nucleotide polymorphisms within the same sequence which were identified as being
CC   located on chromosome 8p12. Also found within the same sequence is the
CC   neuregulin-1-associated gene 1 (NRG1AG1). The gene is useful for treating
CC   schizophrenia in an individual, for diagnosing susceptibility to
CC   schizophrenia, and for screening for agents useful in the treatment of
CC   which is located on chromosome 8p12. Also found within the same sequence
CC   the disease. The present sequence is a fragment of the NRG1AG1 gene of
CC   the invention containing a polymorphic site
XX
SQ   Sequence 401 BP; 98 A; 79 C; 62 G; 161 T; 0 U; 1 Other;

Query Match
Best Local Similarity   2.1%;   Score 167.8;   DB 6;   Length 401;
Matches 220;   Conservative 1;   Mismatches 88;   Indels 0;   Gaps 0;

Qy   3006 TTTCTCACCATCCCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 3065
          |||||||||||||||||||||||||||||||||||||||| ||||||||:||||||
Db   88 TTTGTTACATGGATTATAACTCGGTTCTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy   3066 GTACCCCTTAAGTAATTTCTCACACATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 3125
          |||| ||||| ||||| |||||| ||| ||||||||||| ||||||||||||||
Db   148 GTACCTATTACGTAACTTGTCATTCCTCACCGTTCATTAATATTTAGTCTTTYTGAGTC 207

Qy   3126 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 3185
          |||||| ||||||||||||||| ||||||||||||||||||||||||| |||||
Db   208 TCCAAATGTCTTATTATTCCACCCTCTATGTCCATGTACTCATTAGTTAGTGTCTTTACTTA 267

Qy   3186 TAAGTGATAACATGCAAATATTTGTCTTTCTGTGTGTCTTGTTTTACTTATGAATAATGG 3245
          |||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||
Db   268 TAAGTAAGAACATGCAAATATTTGACTTTCTGTTCTCATCATTATTAGTTGTCTTACTTAG 327

Qy   3246 CCCCCAGTTCTATCTAGGTGTGCTGCAAAAGGCATGAGATTTCATTCTTTTTTTATGGCTATGT 3305
          |||||||||| ||||||| ||||||||||||||||||||||||||||| ||||||||||||
Db   328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTCGAGT 387

Qy   3306 TCTTTCCA 3314
          |||||||
Db   388 AGTATTCCA 396

RESULT 14
AC   ABT02405;
XX
ID   ABT02405 standard; DNA; 401 BP.
XX
AC   ABT02405;
XX
DT   07-NOV-2002 (first entry)
XX
DE   Human neuregulin-1-associated gene 1 fragment SEQ ID NO: 941.
XX
KW   Human; neuregulin 1; neuregulin-1-associated gene 1; NRG1; NRG1AG1;
KW   schizophrenia; chromosome 8p12; single nucleotide polymorphism; SNP;
KW   neuroleptic; gene therapy; splice variant; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US20020094954-A1.
XX
PD   18-JUL-2002.
XX
PF   28-FEB-2001; 2001US-00795686.
XX
PR   28-FEB-2000; 2000US-00515715.
XX

PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR   WPI; 2002-665799/71.
XX
PT   NRG1AG1, useful for treating or diagnosing susceptibility to
PT   schizophrenia, or for assaying a sample for the presence of NRG1AG1
PT   nucleic acid.
XX
PS   Disclosure; Page 599; 700pp; English.
XX
CC   The present invention relates to the human neuregulin-1-associated gene
CC   1 (NRG1AG1), single nucleotide polymorphisms within which were identified
CC   as being associated with an increased susceptibility to schizophrenia,
CC   which is located on chromosome 8p12. Also found within the same sequence
CC   is the neuregulin 1 gene (NRG1). The gene is useful for treating
CC   schizophrenia in an individual, for diagnosing susceptibility to
CC   the disease, and for screening for agents useful in the treatment of
CC   schizophrenia. The present sequence is a fragment of the NRG1AG1 gene of
CC   the invention coontaining a polymorphic site'
XX
SQ   Sequence 401 BP; 98 A; 79 C; 62 G; 161 T; 0 U; 1 Other;

Query Match
Best Local Similarity   2.1%;   Score 167.8;   DB 6;   Length 401;
Matches 220;   Conservative 1;   Mismatches 88;   Indels 0;   Gaps 0;

Qy   3006 TTTCTCACCATCCGTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 3065
          |||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db   88 TTTGTTACATGGATTATAACTCGGTTCTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy   3066 GTACCCCTTAAGTAATTTCTCACACATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 3125
          |||||| ||||| ||||| ||||||||| ||||||||||| ||||||||||||||
Db   148 GTACCTATTACGTAACTTGTCATTCCTCACCGTTCATTAATATTTAGTCTTTYTGAGTC 207

Qy   3126 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 3185
          |||||| ||||||||||||||| |||||||||||||||||||||||||| |||||
Db   208 TCCAAATGTCTTATTATTCCACCCTCTATGTCCATGTACTCATTAGTTAGTGTCTTTACTTA 267

Qy   3186 TAAGTGATAACATGCAAATATTTGTCTTTCTGTGTGTCTTGTTTTACTTATGAATAATGG 3245
          ||||||||||||||| ||||||||||||| |||||||||||||||||||||||||||
Db   268 TAAGTAAGAACATGCAAATATTTGACTTTCTGTTCTCATCATTATTAGTTGTCTTACTTAG 327

Qy   3246 CCCCCAGTTCTATCTAGGTGTGCTGCAAAAGGCATGAGATTTCATTCTTTTTTTATGGCTATGT 3305
          ||||||||| |||||||| |||||||||||||||||||||||||||| ||||||||||||
Db   328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTCGAGT 387

Qy   3306 TCTTTCCA 3314
          ||||||||
Db   388 AGTATTCCA 396

RESULT 15
AC   AAK91687;
XX
ID   AAK91687 standard; cDNA; 847 BP.
XX
AC   AAK91687;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human cDNA 5'-end sequence, SEQ ID NO: 147.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-00114089.
XX

```
PR   08-JUL-1999;  99JP-00194486.
PR   11-JAN-2000;  2000JP-00118774.
PR   02-MAY-2000;  2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI   Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR   WPI; 2001-524255/58.
XX
PT   830 Primers useful for synthesising full length cDNA clones and their use
PT   in genetic manipulation.
XX
PS   Claim 2; SEQ ID NO 147; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA, easily without any special
CC   methods. The present sequence is the nucleotide sequence of the 5'-end of
CC   a cDNA provided in the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in CD-
CC   ROM format directly from EPO.
XX
SQ   Sequence 847 BP; 272 A; 143 C; 199 G; 229 T; 0 U; 4 Other;
```

```
Db      731 CAAAGCATTATATTTTGGGAGTGATTCTTCTTCTTAAGCCACACCA  772
```

```
Search completed: September 11, 2004, 11:15:15
Job time : 1843 secs
```

```
Query Match              2.1%;  Score 166.2;  DB 4;  Length 847;
Best Local Similarity   66.3%;  Pred. No. 2e-28;
Matches  386; Conservative  0; Mismatches  164;  Indels  32;  Gaps  9;

Qy    1245 CAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTACATGACATGGGA 1304
            ||||||||||||||  |  |  ||        |||         |||||||||||||||
Db     213 CAAGTTAATAACAGAAAAGGTATAACAAA-----TTAATCAAGANTTARTGATGACCCAAGA 267

Qy    1305 GTCTTCAGAAATGAAGAACAACCAGGGAAACTGTCCGTGTTTTTTTTTGCTGAGGTT 1364
            |  |||||||||| |||  |  || ||||||||||||||||||||   ||||| ||||
Db     268 GCCTTCAAAAATGAAGACCCAAAAGATGTAGGGAAAGCTATCTGTTTTTCT--CCTAAGTT 325

Qy    1365 CGATGAAGAANTGGAGNTAGCANTGCANTGTAGCCATGTAGNATTAGNACAAAAGGNANTGATCTAGTGGT 1424
            |||  |||||  |||| ||||||||||  |  ||||| ||  |    ||     ||||||
Db     326 TAACGTAGAAAGTGGNACAGCCCGTCTAGAAATGCGACTGGNCANAAAAGGNATATGANTTTGCTGAC 385

Qy    1425 AAAGGNACTCAGGGGAGGAAACACAGCAAAGGCCGTCGTCTATTCAGATTCTTCCTTGATCTCTCTC 1484
            ||||  ||||| ||||| |  |  | | ||    ||||    ||||||||    ||| ||||||
Db     386 AATAGACTGAGAGTGTGGCAAAACCCA--GCCAGGCTGTCCCAGNCACTTCTTGGCTCTCTC-- 442

Qy    1485 TCTCTAATGTATAGACANTTCTTTCCCCTGAGTATGGGGGCAGGACTCTCTTCTTCAAATGAGGGT 1544
            ||||  || |   |    |||||| ||| |||||||| ||||| ||||| |||||||||||||
Db     443 ------TGTGTAACATTTCCTTCCGGGTTATAGGGCACAGCGCCCCACTGGAATGAGGGT 496

Qy    1545 CTTCAAGGGAGAAGGGAGNAAAGTGG------CCTTTTTAGATTTTTATGGCGTTGCTTCG 1596
            ||||| ||||||||||||  |||||       |||  || ||||||| |||| |||||||
Db     497 CTTCAAGGAGAGAAGGGAGAAAGGGAGAGTAACCTTCAGAAGTTTTATGACTTGCTGTG 556

Qy    1597 GGGA-AGAGGAGNTCTAGNTTCTATGACCCATCTTGGGAGAAGGGAATTCTGGTTTCTGT 1655
            |||| |||  |||  || |  ||  ||||| ||||||||||||  ||||||| ||||||||
Db     557 GGAAGAGAGGGGAATTCAAGTGTCTATGACCTGCCTGGAGAAGACGAATTTATTTCCTAT 616

Qy    1656 G-ACTTGCTTCATGAAGAGAAAGAGGAGTAAGAGGGCCAGGAGAAGGTCAGAAAG 1714
            | ||||||||||||||||||||||||||||  ||||||||||||| ||| ||||||
Db     617 GAACTCACTATGGGGGAGAAAGGGGGCCTGGAGAAGACAGGGGAGGAAAG---GGCAG 673

Qy    1715 AGACTTGGCTGCTTCTGAGGGGCTTCCGCGTCTCCTCCTTAGTTCCAAGTACTTCTTAGCATAC 1774
            |||||||||||||||  ||||||| |  || ||  |||||||||||| |||  |||| |||||
Db     674 ACAGANGCTTGCTTGNAGGGCTTCTAANTTCCTTAGTTCAAAG---TCCTTGGGATGC 730

Qy    1775 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 1816
            |||||||| |||| ||||| ||| |  |||| ||||| || ||
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: September 11, 2004, 10:26:38 ; Search time 348 Seconds
                 (without alignments)
                 12822.874 Million cell updates/sec

Title:
Perfect score: 8041
Sequence: 1 caaatcgaagtcctgggcac.........ctaggatctcaggtggatcc 8041

Scoring table: IDENTITY_NUC
                        Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Post-processing: Minimum Match 10%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                    1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                    2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                    3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                    4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                    5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                    6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	146.8	1.8	786431	4	US-09-751-389-3	Sequence 3, Appli
2	141	1.8	41684	4	US-09-536-059-1	Sequence 1, Appli
3	125	1.6	800	4	US-09-016-434-1091	Sequence 1091, Ap
4	123.4	1.5	98844	4	US-09-791-211-10	Sequence 10, Appl
5	121.2	1.5	332000	4	US-10-027-983-11	Sequence 11, Appl
6	119.6	1.5	3726	4	US-09-601-777-1	Sequence 1, Appl
7	119.6	1.5	44848	4	US-09-435-739-42	Sequence 42, Appli
8	119.2	1.5	55827	4	US-09-813-133A-3	Sequence 3, Appli
9	118.8	1.5	786431	4	US-09-751-389-3	Sequence 3, Appli
10	118	1.5	168575	4	US-09-426-290-1	Sequence 1, Appli
11	118	1.5	168575	4	US-09-426-290-1	Sequence 1, Appli
12	117.6	1.5	168575	4	US-09-426-290-1	Sequence 1, Appli
13	117.4	1.5	6464	4	US-08-321-478-2	Sequence 2, Appli
14	117	1.5	6464	4	US-08-321-478-2	Sequence 2, Appli
15	115.8	1.4	12385	4	US-08-321-862-3	Sequence 3, Appli
16	113	1.4	467	4	US-09-822-862-3	Sequence 3, Appli
17	112	1.4	487	4	US-09-621-976-1582	Sequence 1582, Ap
18	111.6	1.4	116592	4	US-09-621-976-19221	Sequence 19221, Ap
19	111.6	1.4	202001	4	US-09-818-512-3	Sequence 3, Appli
20	111.2	1.4	694	1	US-09-734-674-3	Sequence 3, Appli
21	111.2	1.4	694	3	US-08-358-171-25	Sequence 25, Appl
22	111.2	1.4	694	3	US-09-090-947-25	Sequence 25, Appl
23	110.2	1.4	5000	4	US-09-146-053-3	Sequence 3, Appl
24	109.8	1.4	197996	4	US-09-877-177A-10	Sequence 10, Appl
25	108.8	1.4	116592	4	US-09-818-512-3	Sequence 3, Appli
26	108.4	1.3	90541	4	US-09-759-359A-3	Sequence 3, Appli
27	108.4	1.3	319608	4	US-09-679-409-1	Sequence 1, Appl
28	105.2	1.3	112132	4	US-09-741-150-3	Sequence 3, Appli
29	105.2	1.3	112132	4	US-10-160-187-3	Sequence 3, Appli
30	104	1.3	12047	2	US-09-022-461-1	Sequence 1, Appl
31	104	1.3	12047	4	US-09-033-556-3	Sequence 3, Appli
32	104	1.3	12047	4	US-09-474-699-11	Sequence 11, Appl
33	104	1.3	12047	4	US-09-151-376-3	Sequence 3, Appli
34	103.2	1.3	99916	4	US-09-816-095-3	Sequence 3, Appli
35	102.2	1.3	72604	4	US-09-268-992-7	Sequence 7, Appli
36	102.2	1.3	72604	4	US-09-657-474-7	Sequence 7, Appli
37	101.6	1.3	84495	4	US-09-757-906-3	Sequence 3, Appli
38	101	1.3	4752	1	US-08-201-697-3	Sequence 3, Appli
39	101	1.3	4799	1	US-08-201-697-6	Sequence 6, Appli
40	101	1.3	4815	1	US-08-201-697-5	Sequence 5, Appli
41	99.4	1.2	128779	4	US-09-497-85bA-38	Sequence 38, Appl
42	99.2	1.2	506	1	US-09-621-976-2426	Sequence 2426, Ap
43	99.2	1.2	1426	1	US-08-358-171-3	Sequence 3, Appli
44	98.8	1.2	1426	1	US-09-090-947-3	Sequence 3, Appli
45	98.8	1.2	41100	4	US-09-755-665-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-09-751-389-3/c
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match   1.8%;   Score 146.8; DB 4; Length 786431;
Best Local Similarity   75.6%;   Pred. No. 3.9e-28;
Matches 195; Conservative 0; Mismatches 62; Indels 1; Gaps 1;

Qy   3001 AGTAATTTCTCCACCATCCGTCGACTTCTTTGCCCCCTGGGTTATTCATTCACCTGAATGAGTGAAGT 3060
             |||| ||||||| ||| |||| ||||||||| ||||| || ||||| | ||||| || || |||||
Db   170144 ACTAATATATTGCATATTGGTGAAGTCAGGGCTTTAGTGTACTCATCACCGGAATAAAGT 170085

Qy   3061 GCATTGTACCCCTTAAGTAAATTTCACACATCCGCTGACTTCTTGCCCCTCATCCTTCT 3120
             |||||| ||||| |||||||| ||| || || |||||||||||||| ||| ||| |||
Db   170084 ACAATTGCA-CCATTAAGTAATTTCATCATCCCACCCCCCTTTTACCCTTTCACCTTTCC 170026

Qy   3121 GAGGCTCCATTGTCCATCATTCCACACTCCTACATCTATGTAGTACACAATTATTTTAGCTCCT 3180
             |||||||||||||| |||||||||| || |||||||||| |||| ||||| | |||| |||
Db   170025 AAGTCTCCATTGTCTATCATTCCACAGTCGTTACATCTATGTTGTGCACATTATTTAGCTCTCC 169966

Qy   3181 ACTTATAAGTGATAACATGCAAATATTTGCTTTTCTGTGTCTGTCTTGTTTTTACTTAATGAT 3240
             |||||| ||| || || ||| | ||||||||| |||| |||||| |||||||| ||| |||
Db   169965 ACTTATAAAATGAGAACATGCAGTATTGCGTAATTTGTCTTTCTATGTGTCAGTTTCATTAAGAT 169906

Qy   3241 AATGGCCCCCAGTTTCTAT 3258
             ||||| ||||||||||| |
Db   169905 AATAGCCTGCAGTTCCAT 169888

RESULT 2

~ US-09-536-059-1
~ Sequence 1, Application US/09536059
~ Patent No. 654737
~ GENERAL INFORMATION:
~ APPLICANT: Blumenfeld, Marta
~ APPLICANT: Chumakov, Ilya
~ APPLICANT: Bougueleret, Lydie
~ APPLICANT: Cohen-Akenine, Annick
~ TITLE OF INVENTION: GENOMIC SEQUENCE OF THE purH GENE AND purH-RELATED BIALLELIC
~ TITLE OF INVENTION: MARKERS.
~ FILE REFERENCE: GENSET 058AUS
~ CURRENT APPLICATION NUMBER: US/09/536,059
~ CURRENT FILING DATE: 2000-03-31
~ PRIOR APPLICATION NUMBER: US 60/125,961
~ PRIOR FILING DATE: 1999-03-24
~ NUMBER OF SEQ ID NOS: 24
~ SOFTWARE: Patent_pm
~ SEQ ID NO 1
~ LENGTH: 41684
~ TYPE: DNA
~ ORGANISM: Homo sapiens
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: 1..2000
~ OTHER INFORMATION: 5'regulatory region
~ NAME/KEY: exon
~ LOCATION: 2001..2096
~ OTHER INFORMATION: exon 1
~ NAME/KEY: exon
~ LOCATION: 2433..2559
~ OTHER INFORMATION: exon 2
~ NAME/KEY: exon
~ LOCATION: 8092..8168
~ OTHER INFORMATION: exon 3
~ NAME/KEY: exon
~ LOCATION: 9600..9666
~ OTHER INFORMATION: exon 4
~ NAME/KEY: exon
~ LOCATION: 15178..15266
~ OTHER INFORMATION: exon 5
~ NAME/KEY: exon
~ LOCATION: 15924..16075
~ OTHER INFORMATION: exon 6
~ NAME/KEY: exon
~ LOCATION: 16759..16915
~ OTHER INFORMATION: exon 7
~ NAME/KEY: exon
~ LOCATION: 22309..22434
~ OTHER INFORMATION: exon 8
~ NAME/KEY: exon
~ LOCATION: 23277..23384
~ OTHER INFORMATION: exon 9
~ NAME/KEY: exon
~ LOCATION: 24841..24926
~ OTHER INFORMATION: exon 10
~ NAME/KEY: exon
~ LOCATION: 25957..26046
~ OTHER INFORMATION: exon 11
~ NAME/KEY: exon
~ LOCATION: 28700..28828
~ OTHER INFORMATION: exon 12
~ NAME/KEY: exon
~ LOCATION: 34699..34791
~ OTHER INFORMATION: exon 13
~ NAME/KEY: exon
~ LOCATION: 36679..36881
~ OTHER INFORMATION: exon 14
~ NAME/KEY: exon
~ LOCATION: 39014..39169
~ OTHER INFORMATION: exon 15
~ NAME/KEY: exon
~ LOCATION: 39456..39684
~ OTHER INFORMATION: exon 16

~ NAME/KEY: misc_feature
~ LOCATION: 39685..41684
~ OTHER INFORMATION: 3'regulatory region
~ NAME/KEY: allele
~ LOCATION: 6491
~ OTHER INFORMATION: 99-32284-107 : polymorphic base C or T
~ NAME/KEY: allele
~ LOCATION: 15234
~ OTHER INFORMATION: 99-5602-372   : polymorphic base G or C
~ NAME/KEY: allele
~ LOCATION: 15868
~ OTHER INFORMATION: 99-22573-321 : polymorphic base C or T
~ NAME/KEY: allele
~ LOCATION: 16729
~ OTHER INFORMATION: 5-290-32   : polymorphic base C or T
~ NAME/KEY: allele
~ LOCATION: 18572
~ OTHER INFORMATION: 99-22586-39   : polymorphic base C or T
~ NAME/KEY: allele
~ LOCATION: 22906
~ OTHER INFORMATION: 99-5596-197   : polymorphic base A or G
~ NAME/KEY: allele
~ LOCATION: 23175
~ OTHER INFORMATION: 5-293-76   : polymorphic base C or T
~ NAME/KEY: allele
~ LOCATION: 23253
~ OTHER INFORMATION: 5-293-155   : polymorphic base A or G
~ NAME/KEY: allele
~ LOCATION: 26106
~ OTHER INFORMATION: 99-23454-105 : polymorphic base G or G
~ NAME/KEY: allele
~ LOCATION: 31250
~ OTHER INFORMATION: 99-15528-333 : polymorphic base A or G
~ NAME/KEY: allele
~ LOCATION: 30464
~ OTHER INFORMATION: 5-294-285   : polymorphic base G or G
~ NAME/KEY: allele
~ LOCATION: 35148
~ OTHER INFORMATION: 99-23454-317 : polymorphic base A or G
~ NAME/KEY: allele
~ LOCATION: 30669
~ OTHER INFORMATION: 99-22586-39   : polymorphic base A or G
~ NAME/KEY: allele
~ LOCATION: 37286
~ OTHER INFORMATION: 99-15798-86   : polymorphic base A or G
~ NAME/KEY: allele
~ LOCATION: 36801
~ OTHER INFORMATION: 5-297-209   : polymorphic base A or G
~ NAME/KEY: allele
~ LOCATION: 39669
~ OTHER INFORMATION: 99-23460-199 : polymorphic base G or T
~ NAME/KEY: allele
~ LOCATION: 37536
~ OTHER INFORMATION: 99-32281-276 : polymorphic base C or T
~ NAME/KEY: allele
~ LOCATION: 39331
~ OTHER INFORMATION: 99-32281-26   : polymorphic base C or T
~ NAME/KEY: primer_bind
~ LOCATION: 6577..6597
~ OTHER INFORMATION: 99-32284.pu complement
~ NAME/KEY: primer_bind
~ LOCATION: 6137..6157
~ OTHER INFORMATION: 99-32284.rp
~ NAME/KEY: primer_bind
~ LOCATION: 14864..14882
~ OTHER INFORMATION: 99-5602.pu
~ NAME/KEY: primer_bind
~ LOCATION: 15292..15312
~ OTHER INFORMATION: 99-5602.rp complement
~ NAME/KEY: allele
~ LOCATION: 39669
~ OTHER INFORMATION: 5-298-376   : polymorphic base A or G

```
~; LOCATION: 15837..15855
~; OTHER INFORMATION: 5-290.pu
~; NAME/KEY: primer_bind
~; LOCATION: 16249..16266
~; OTHER INFORMATION: 5-290.rp complement
~; NAME/KEY: primer_bind
~; LOCATION: 16599..16617
~; OTHER INFORMATION: 99-22573.rp
~; NAME/KEY: primer_bind
~; LOCATION: 17030..17049
~; OTHER INFORMATION: 99-22573.pu complement
~; NAME/KEY: primer_bind
~; LOCATION: 18131..18150
~; OTHER INFORMATION: 99-22586.rp
~; NAME/KEY: primer_bind
~; LOCATION: 18592..18610
~; OTHER INFORMATION: 99-22586.pu complement
~; NAME/KEY: primer_bind
~; LOCATION: 22710..22727
~; OTHER INFORMATION: 5-294.pu
~; NAME/KEY: primer_bind
~; LOCATION: 23100..23118
~; OTHER INFORMATION: 5-294.rp complement
~; NAME/KEY: primer_bind
~; LOCATION: 23130..23149
~; OTHER INFORMATION: 5-293.pu
~; NAME/KEY: primer_bind
~; LOCATION: 23512..23530
~; OTHER INFORMATION: 5-293.rp complement
~; NAME/KEY: primer_bind
~; LOCATION: 25822..25840
~; OTHER INFORMATION: 99-5596.rp complement
~; NAME/KEY: primer_bind
~; LOCATION: 26222..26241
~; OTHER INFORMATION: 99-5596.pu
~; NAME/KEY: primer_bind
~; LOCATION: 30332..30352
~; OTHER INFORMATION: 99-23454.rp complement
~; NAME/KEY: primer_bind
~; LOCATION: 30754..30773
~; OTHER INFORMATION: 99-23454.pu complement
~; NAME/KEY: primer_bind
~; LOCATION: 30918..30935
~; OTHER INFORMATION: 99-15528.pu
~; NAME/KEY: primer_bind
~; LOCATION: 31390..31408
~; OTHER INFORMATION: 99-15528.rp complement
~; NAME/KEY: primer_bind
~; LOCATION: 34780..34799
~; OTHER INFORMATION: 99-15798.rp
~; NAME/KEY: primer_bind
~; LOCATION: 35215..35233
~; OTHER INFORMATION: 99-15798.pu complement
~; NAME/KEY: primer_bind
~; LOCATION: 36593..36610
```

```
Query Match          1.8%;  Score 141;  DB 4;  Length 41684;
Best Local Similarity  63.7%;  Pred. No. 2.1e-27;
Matches  375;  Conservative  175;  Mismatches  0;  Indels  39;  Gaps  9;

Qy  1162  CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCCA  1221
               |  ||||||||| |||||||||||   |||||||  ||||| |||||||||  |||
Db  4727  CCAAGATTTTGTAGGGGTTGGAAAAGTGTGGATCC--TTTGCTTCTCCATCGTAAACGTCACG  4784

Qy  1222  ACCAAAACTCCTATAGTAAAGAGACAAGGT--TAATAAGAGCAAAAACCTAACAAAATTTATTT  1279
          |||||||||||||||||||   ||||||||   |||||||||||||||||  |||||| ||||
Db  4785  GCCAAATATTTTTATAGAGAGAAGAGACAGGTTATAATAAGAGAAAAGCCATAACAAAATTTATTT  4844

Qy  1280  AATCAAAAGTTTTACATGACATGGGGAGTCTTCAGAAATGAAGACCCAAGACCCAGGGGGAA  1339
          |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db  4845  AA-CAAAGTTTTACATGACATGAAGAGCCTTCAG-AATGAAGACCCAGAGACCCAGAGGGAAA  4902

Qy  1340  ACTGTCTGTGTTTTTTTTGCTGAGGTTCGATGAATGAAGAAGAATGACATGTAGCCATGTAGAGTT  1399
```

RESULT 3
US-09-016-434-1091
; Sequence 1091, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1091:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK

```
Db  4903  ACCATC-CATTTTTATGTTTAGGTTCAACAAAGAATGACAAAGAATGGAAGTATGATT  4961
          ||||||  ||||| ||||| |||| ||||  |||| |||| |||  ||  |||| ||||
Qy  1400  AGACAAAAGGATATGATCTAGTGGTTAAAGGACTCAG--GGGGAAAACACAGCAAGGCCTGT  1457

Db  4962  GGACAGCAAGGAATATGGTGTATGCTAGTAGACTCAGG-TGGAGAAAAGCCAGGAAAGCCTGT  5021
          |||| | |||| | |||||| |||  |||| |||||||  |||||||||| ||||||||||||
Qy  1458  CTATTCAGATTCTTCTTGATCTCTCTCTCTATGT------ATAGCATTCTTTCCT  1508

Db  5022  CTGTCCAGATTCTTCTTGGCTTCTCTAAAATTCTTTCTCCACCCCCTTAGGGATCTCCTGA  5081
          |||| |||||||||||| | ||||||    ||| ||||||     |||| | |  |||| |
Qy  1509  CCTGAGTATGGGGCAGGACTCTTCTTCATGAGGGTCTTCAAGGGAGGAAGGGAGGAAAG--  1566

Db  5082  CCTACTAAATGGGGCAAGGAAGATGAGGAGGGATTTCTTTACGTTCCAGCTCCTTAGACAGAAAGCC  5141
                 ||||||||| | ||  ||| | ||| ||||| |||| ||||| |  ||||| |   ||||||
Qy  1567  -----------TGGCTTTTTAGATTTTATGGGCTTTGCTTCGGGGAGGAGGGGAAGAGGAGAGTTCT  1611

Db  5142  AGTGGAAAGTTAGAATCATAAGTTTAAATCTTATGACTGGCTTTTACGTTCGTTGACTTGCTTTCATCGA  5201
          |||  ||||||| |||| | |||||||| |||| || |||| ||| || ||| |||| ||||||
Qy  1612  AGTTTCTATGACCCATCTTGGGGAGGAGGAATTCTGGTTTCGTGTGACTTGTTTCGTTGCTTTCATGAA  1671

Db  5202  AGTTCTGTAAACTGCCCTGGGGAAGAGAAAAATTCTCATTTCTGTGACTTCAGG----GAA  5257
          |||| || ||| |||| |||||||||| |||| ||| |||| ||| | ||||    |||
Qy  1672  GAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGGAGGATGGGTCAGAAAGAGACTTT  1720

Db  5258  GAATGAAGGGTGAGGAGGCAAGAGGGCAGGAGAGGAAAGCTATATATATATTT  5306
```

CLONE: 91216372
US-09-016-434-1091

Query Match   1.6%; Score 125; DB 4; Length 800;
Best Local Similarity 100.0%; Pred. No. 3.7e-24;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  6364 GCAAGTACGCTCAAAGCCGGAAGCCACAGCTCCTCCTTCCCGCCGCCAATTCTTCTTCCTTGCTTGCG 6423
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17 GCAAAGTACGCTCAAAGCCGGAAGCCACAGCTCCTCCTTCCCGCCGCCAATTCTTCTTCCTTGCTTGCG 76

Qy  6424 AATTCCAAGCTGTTAAATAAGATGATGTGCAAAGGGGTTGCAGGTTGCCGGCTTGCTTCTTGCTTG 6483
         |||||||||||||||||||||||| ||||||||||| ||||||| ||||||||||| ||||||||||
Db    77 AATTCCAAGCTGTTAAATAAGATGTGCAAAGGGGCTTGCAGGTCTGGCCGGCTTCTTGCTTG 136

Qy  6484 AGGAG 6488
         |||||
Db   137 AGGAG 141
```

RESULT 4
US-09-791-211-10
Sequence 10, Application US/09791211
Patent No. 6448080
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
APPLICANT: Andrew T. Watt
TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
FILE REFERENCE: RTS-0205
CURRENT APPLICATION NUMBER: US/09/791,211
CURRENT FILING DATE: 2001-02-23
NUMBER OF SEQ ID NOS: 90
SEQ ID NO 10
LENGTH: 98844
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 24962
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 64383
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 65468
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 65469
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 65470
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 65471
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 87130
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 89049
OTHER INFORMATION: unknown

US-09-791-211-10

Query Match   1.5%; Score 123.4; DB 4; Length 98844;
Best Local Similarity 70.0%; Pred. No. 2.7e-22;
Matches 166; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy  3107 CCCCTCATCCTTCTGAGGCTCCATGTCATCCATCTTACACTCTATGTGTACAC 3166
         || ||||| | || ||  |  || ||  | |    | ||||||| ||||  |
Db 16525 CCACCCATCCATCCCAGCCCTCTGGTTAATCACCAATTCTACTTACTTACCTCAATGACAATCAA 16584
```

```
Db 16585 ATTTTTTAGCTCCCACATATPAAGTGAGAGAACATGCGGATATTTGTCTTTCCCGTCGTTGGTTT 16644
```

```
Qy  3227 GTTTTACTTATGATAATGCCCCAGTTCTTATCTAGGCTGCTGCAAAGGCATGATTTCA 3286
         |||| ||| || | | || ||||||| |  ||  | ||||||||||| | |||| |||
Db 16645 ATTTCACTTAACAGAATGTCTTCAGTTCCATTCATGTTGCTGCAAATGACAGGAGTTCA 16704

Qy  3287 TTCTTTTTTATGGGTATAGTTCTTTCCCCAATTTAGATAAAGAGAACACTCGGCACTTGCTC 3343
         ||||||  | |||| |  | | ||| ||  | | |  ||  | || |  || |||||  |||
Db 16705 TTCTTTATTATGGCGTTAATACTATTCCATTGTATGTATGTATACACATTTTCTTTATC 16761
```

RESULT 5
US-10-027-983-11/c
Sequence 11, Application US/10027983
Patent No. 6617162
GENERAL INFORMATION:
APPLICANT: Mark P. Roach
APPLICANT: Kenneth W. Dobie
TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
FILE REFERENCE: RTS-0340
CURRENT APPLICATION NUMBER: US/10/027,983
CURRENT FILING DATE: 2001-12-18
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 11
LENGTH: 392000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 137740
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 137742
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 145507
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 151967
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (138122)...(138221)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc_feature
LOCATION: (151967)...(1542066)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 154217
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (164037)...(164136)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc_feature
LOCATION: (174657)...(174756)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc_feature
LOCATION: (186224)...(186323)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc_feature
LOCATION: (195542)...(195641)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 202703
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (202771)...(202870)
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (206246)...(215602)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc_feature
LOCATION: (218176)...(218225)

```
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: misc_feature
;  LOCATION: (220560)...(220459)
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: misc_feature
;  LOCATION: (222717)...(222816)
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: misc_feature
;  LOCATION: (223981)...(224080)
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: misc_feature
;  LOCATION: (227487)...(227586)
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: misc_feature
;  LOCATION: (230157)...(230256)
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: misc_feature
;  LOCATION: (232299)...(232398)
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: misc_feature
;  LOCATION: (236552)...(2366651)
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: misc_feature
;  LOCATION: (238789)...(248788)
;  OTHER INFORMATION: n = A,T,C or G
;  NAME/KEY: exon
;  LOCATION: (118288)...(119101)
;  OTHER INFORMATION: exon 1C
;  NAME/KEY: exon:intron junction
;  LOCATION: (151129)...(151130)
;  OTHER INFORMATION: exon 5:intron 5
;  NAME/KEY: exon:intron junction
;  LOCATION: (292248)...(292249)
;  OTHER INFORMATION: exon 9:intron 9
;  NAME/KEY: exon:intron junction
;  LOCATION: (348578)...(348579)
;  OTHER INFORMATION: exon 10:intron 10
;  NAME/KEY: intron
;  LOCATION: (348579)...(381838)
;  OTHER INFORMATION: intron 10
;  NAME/KEY: intron:exon junction
;  LOCATION: (386185)...(386186)
;  OTHER INFORMATION: intron 11:exon 12
US-10-027-983-11

Query Match             1.5%;  Score 121.2;  DB 4;  Length 392000;
Best Local Similarity   66.8%;  Pred. No. 2.9e-21;
Matches 203;  Conservative  0;  Mismatches  98;  Indels   3;  Gaps   2;

Qy     3042 TTCATCACCTGAATGATGTGCATTGTACCCTTAAGTAATTTCTCACCATCCGCTGACTT 3101
            ||||||||||||||||||||||||| |||||||||||||||||| ||||||||||||||
Db    99765 TTCATCACCCAGGTATTAAGCCTAGTACCCATT-AGTTATTTTTCCTGATCCTCTCCCTC 99707

Qy     3102 CTTGCCCCCTCATCCTTCTG--AGGCTCCATTGTCCATCATTCCACACTCTACATCTATG 3159
            ||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db    99706 CTTCCACCCTCCACCCTCGATAGGCCCCAGTGTCTGTTATTCCTCTCTATGTGTCCCATG 99647

Qy     3160 TGTACACATTATTTAGCTCCTACTATAAGTGATAACARCGCAATATATTTGTCTTTCTGTGT 3219
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3219 AATTCTAATCATGTAGCTCCCACTTACACAAGTGAGGAACAATGCAGTATTAGGTTTTTCTGTTC 99587

Qy     3220 CTGTCTTGTTTTACTTATTATGATAATGGGCCCCCCAGTTCATTCATCTAGGCTGCTGCAAAAGGCAT 3279
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    99586 CTGCATTAGTTTGCTAAGGATAANTGGCCTTCCAGCTCCTTCCAGTTCCTGAAAAGGACAT 99527

Qy     3280 GATTTCATTTCTTTTTTATGGCTACTGTTCTTTTCCAATTAGATAAAAGAACACTCGCACTT 3339
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3280 GATCCTCATCTCTTTTTGGTGGCTGCATAGTAATTCATCAGTGTATAGTGTACCACACATTTTCTT 99467

Qy     3340 GCTC 3343
            ||||
Db    99466 TATC 99463
```

```
RESULT 6
US-09-601-777-1
;  Sequence 1, Application US/09601777
;  Patent No. 6461848
;  GENERAL INFORMATION:
;  APPLICANT: Nakajima, Motowo
;  APPLICANT: Funakubo, Minako
;  TITLE OF INVENTION: Human heparanase polypeptide and cDNA
;  FILE REFERENCE: 30384A
;  CURRENT APPLICATION NUMBER: US/09/601,777
;  CURRENT FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
;
;  SEQ ID NO 1
;  LENGTH: 3726
;  TYPE: DNA
;  ORGANISM: Human
US-09-601-777-1

Query Match             1.5%;  Score 119.6;  DB 4;  Length 3726;
Best Local Similarity   65.4%;  Pred. No. 3.2e-22;
Matches 208;  Conservative  0;  Mismatches 104;  Indels   6;  Gaps   2;

Qy     3039 GTATTCATCACCTGAATGATGTGCATTGTACCCTTAAGTAATTTCTCACCATCCGCTGA 3098
            ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     3233 GTACTTGTCACCCGAGTGATGTACATTGTACCCAATAGGTAATTTTTCATCACATCTAT 3292

Qy     3099 CTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTAT 3158
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3293 CCTTCCGCCCGCTCTTC-CCTTCTCGAGTCTCCACCTGTCTGTTATTCCTCTCTATGTGT 3351

Qy     3159 GTGTACACATTATTTAGCTCCTACTTATATAAGTGATAACATGCAATATATTTTGTCTTTCTGTG 3218
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3293 CCTCCGCCCGCTCTTC-CCTTCTCGAGTCTCCACCTGTCTGTTATTCCTCTCTATGTGT 3351

Qy     3219 TCTGTCTTGTTTTACTTATGATAATGGGCCCCCAGTTCATTCATCTAGGCTGCTGCAAAAGGCA 3278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3412 CCTGAGTTACTTCCCTAGGATAAACAGCCCCCCAGTTCGTCGAAAGTTGCTGAAAATACA 3471

Qy     3279 TGATTTCATTTCTTTTTTTATGGCTAGTGTTCTTTTCCAATTAGATAAAAGAACACTCGCACT 3338
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3412 CCTGAGTTACTTCCCTAGGATAAACAGCCCCCCAGTTCGTCGAAAGTTGCTGAAAATACA 3471

Qy     3339 TGCTCTTACTTCTAGTTG 3356
            ||  ||||||||||||||
Db     3527 TTAATCCACTTATCAGTTG 3544
```

```
RESULT 7
US-09-435-739-42
;  Sequence 42, Application US/09435739
;  Patent No. 6664105
;  GENERAL INFORMATION:
;  APPLICANT: Pecker, Iris
;  APPLICANT: Vlodavsky, Israel
;  APPLICANT: Feinstein, Elena
;  TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE HAVING HEPARANASE ACTIVITY
;  TITLE OF INVENTION: EXPRESSION OF SAME IN GENETICALLY MODIFIED CELLS
;  FILE REFERENCE: 00/20454
;  CURRENT APPLICATION NUMBER: US/09/435,739
;  CURRENT FILING DATE: 2001-06-05
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: PatentIn version 3.0
;
;  SEQ ID NO 42
;  LENGTH: 44848
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-435-739-42

Query Match             1.5%;  Score 119.6;  DB 4;  Length 44848;
Best Local Similarity   65.4%;  Pred. No. 1.8e-21;
Matches 208;  Conservative  0;  Mismatches 104;  Indels   6;  Gaps   2;
```

RESULT 8
US-09-813-133A-3/c
Sequence 3, Application US/09813133A
Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-3

Query Match   1.5%; Score 119.2; DB 4; Length 55827;
Best Local Similarity   62.6%; Pred. No. 2.7e-21;
Matches 218; Conservative 0; Mismatches 127; Indels   3; Gaps   2;

Qy   3039 GTATTCATCACCTGAATGATGATGTGCATTGTACCCTTAAGTAATTTCTCACCATCCGCTGA 3098
            |||||||||||||||| | | | ||||||||||||||| ||||||| |||||||||
Db   44272 GTACTTGTCACCTGAGTGAGTGACATTGTACCAATAGGTAATTTTTCATCCAATCCACCCT 44331

Qy   3099 CTTCTTGCCCCCTCATCCTTCTCGAGGCTCCATTGTCCATCATTCCACACTCTACATCTAT 3158
            ||| | |||||| ||||| ||| || ||||| |||| ||||| ||||| || | |||||
Db   44332 CCTTCCGCCCCTCTC--CCTTCTGAGTCTCCAACATCCCTATTACCACTGTGTTACTCTT 44390

Qy   3159 GTGTACACATTATTTAGCTCCTACTTATATAGTCAATATAGTCGAATAGGTGAGAACATGA 3218
            | | |||| |||||||| ||| |||| | ||| ||| || | ||| | ||| | ||
Db   44391 GTGTACCTACAGCGTAAGCTCACACTTATAAGTGAGAAGGAGAACAGTGCAGTGATTTCCATT 44450

Qy   3279 TGATTCATTCTTTTTTATTGGGCTATGTTCTTTTCCCAATTTAGATAAAGAACACTCCGACT 3338
            |||| || || ||| |||| | ||| || || | |||| | | | | || | || ||
Db   44511 ----TTATTCTTCTTATGGCTGAGTAATAAGTCATGGTACATATATACCACACATTTCT 44565

Qy   3339 TGCTCTTTACTTCTATTG 3356
            ||||||| | | |||||
Db   44566 TTATCCACTATCAGTTG 44583

RESULT 9
US-09-751-389-3
Sequence 3, Application US/09751389
Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match   1.5%; Score 118.8; DB 4; Length 786431;
Best Local Similarity   65.6%; Pred. No. 2.1e-20;
Matches 193; Conservative 0; Mismatches 92; Indels   9; Gaps   1;

Qy   3037 GGGTATTCATCACCTGAATGATGATGTGCATTGTACCCTTAAGTAATTTCTCACCATCCGCT 3096
            || ||||||||||||||| | | | ||||||||||||| ||||||| |||||||||
Db   320220 GTGCACCCATCACCTGAGTAGTGCCCCTGTACCCAGTAGTCACTGAGTACTTTCATCATCCCTCTT 320279

Qy   3097 GACTTCTTGCCCCCTCATCCTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 3156
            || ||||| |||||| ||||| |||| ||||| |||| ||||| ||||| | | |||
Db   320280 AACCTC-----CCACTTCGAGTCTCCGGTGTCCATTATACCACTCTATTTACCATCTCT 320330

Qy   3157 ATGTGTACACATTATTTAGCTCCTACTTATATAGTGAATAACATGCAATATTGTCTTTCTG 3216
            |||||||| ||| |||||| ||| |||| | | | || || |||| | || ||| ||
Db   320331 TTGCATACCCTAGCTCAGCTCCTACTTACAAGTGAGACAGCAGTATTTGGTTTTTGA 320390

Qy   3217 TGTCTGTCTTGTTTTACTTAGATAAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 3276
            ||| ||| |||| |||| ||| || | ||||||||||| |||| ||| ||| ||||||
Db   320391 TTCCTGAGTACTTCACTTAAAATAAATGTCTCTCCAGTACCACCCAAGTTGCTGCAAAAGG 320450

Qy   3277 CATCAATTCATTCTTTTTTATGCTTTTTCCCAAATTTAGATAAAGAACA 3330
            ||| ||| |||||||||||||| ||| | |||||| || | ||||||||
Db   320451 CATGAATTTTATTCTTTTTGTGGCTGAGTAGGAATTCTATGGTATATATATGGTATCA 320504

RESULT 10
US-09-426-290-1
Sequence 1, Application US/09426290
Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)

Qy   3039 GTATTCATCACCTGAATGATGATGTGCATTGTACCCTTAAGTAATTTCTCA--CCATCCGCT 3096
            |||||||||||||||| | | | ||||||||||||| ||||||| ||||||| ||||||||
Db   43777 GAACCCATCACCCAAATGGTGAACACAGTACCCACTTAGTAGTACTTTTCAATCTCTTCCCC 43718

Qy   3097 GACTTCTTGCCCCCTCATCCTTCTCGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 3156
            || ||||| |||||| |||| |||| || ||||| |||| ||||| ||||| || | |||
Db   43717 TCTCTCCTTCCTCCGCCCCCTTTTTGACGTCCCAGTGTCT-ATTCCCATCTTTAGTCT 43659

Qy   3157 ATGTGTACACATTATTTAGCTCCTACTTATATAAGTGATGATAACATGCAATATTTGTCTTTCTG 3216
            |||||||| ||| |||||| ||| |||| | | | || |||| ||| | | || ||| ||
Db   43658 GTGTATACTCAGTGTTAGCTTTTAACTTATTAGAACCAGTATTTGTCTTTCTG 43599

Qy   3217 TGTCTGTTTTGTTTTTACTTATGATAATGGGCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 3276
            ||| ||| |||| |||| || ||| | | ||||||||||| |||| ||| ||| ||||||
Db   43598 TTTCTGTGTTAATTTACTTAGGACAAATGCCTTCCAGCTGCCATCCAGTCTGCGTAAAGGA 43539

Qy   3277 CATGAATTTCATTTCTTTTTTATGCTTTTTCCCAAATTTAGATAAAGAACA 3336
            ||| ||| |||||||||||||| ||| || |||||| || | |||||||| |
Db   43538 CATAATTTGTTCTTTTTTATGGGCTGTGTTGTAGTAATCCCAGTGGTATATTTACACACATTTT 43479

Qy   3337 CTTGCTCTTTACTTCTATTGGAATACTAATTCCTAGGGCTTCTTTGCATT 3384
            || ||||| |||| ||| ||| | |||| | ||| | | | | ||
Db   43478 CTTTATCCAATCCGCATTGGACAGACACCTTAGAATTGATTCCATGTCTT 43431

NAME/KEY: CDS
LOCATION: (101753)...(101996)
NAME/KEY: CDS
LOCATION: (110324)...(110439)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)

US-09-426-290-1

Query Match   1.5%; Score 118; DB 4; Length 168575;
Best Local Similarity 72.1%; Conservative 0; Mismatches 70; Indels 6; Gaps 3;
Matches 196; Pred. No. 1.2e-20;

Qy   3039 GTATTCATCACCTGAATGAATGGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGA 3098
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142678 GTAGTCATCAATCTGAACAGTGAAC-TTGTACCAAATAAGTAATTTTCAACTTCTC---AT 142733

Qy   3099 CTTCTTGCCCCTCCATCCTTCTGAGGCGTTCCATTGTCATCATTCCACACTTCACATCTAT 3158
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142734 CCACCCACCCTCCCAATCTTTTGTAGTACCCAAGGTCTTCATTAT--CCCACTCTGTATGCCT 142791

Qy   3159 GTGTACACATTATTAGCTCCCTACTTATTAGTGATTACATGCAATATTTTGTCTTTCGTTG 3218
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142792 GTGTACCTATTGTTTAGCTTCCACTTATAAGTGAACACATGCAGCATTGACTTTCTGTT 142851

Qy   3219 TCTGTCTTGTTTTACTTATANGATAATAATGAGCCCCCCCAGTTCTATCTNCTGGCTCGCA 3278
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142852 TCTGAGTTATTTACTTAGGATAATGGCCTCCAGTTCCATCCTTACATGCTGCAAAAGTTA 142911

Qy   3279 TGATTTCATTCTTTTTATTTATGGCTATGTTCTTT 3310
            |||||||||||||||||||||||||||||||||||
Db 142912 TGATTTATTCTTTTTTATGGCTCCATTATAT 142943

RESULT 11
US-09-426-290-1/c
Sequence 1, Application US/09426290
Patent No. 6410712
GENERAL INFORMATION:
APPLICANT: Berglind Ran Olafsdottir
APPLICANT: Jeffrey Gulcher
TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
FILE REFERENCE: 2345.2001-000
CURRENT APPLICATION NUMBER: US/09/426,290
CURRENT FILING DATE: 1999-10-25
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 168575
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (21181)...(21403)
NAME/KEY: CDS
LOCATION: (95252)...(95430)
NAME/KEY: CDS
LOCATION: (101753)...(101996)
NAME/KEY: CDS
LOCATION: (110324)...(110439)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)

US-09-426-290-1
Query Match   1.5%; Score 117.6; DB 4; Length 168575;
Best Local Similarity 66.0%; Pred. No. 1.6e-20;

Matches 186; Conservative 0; Mismatches 94; Indels 2; Gaps 1;

Qy   3077 GTAATTTCTCACCATCCGCTGACTTCTTGCCCCTCCATCCTTCTG--AGGCTCCATTGTC 3134
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 100536 GTTATTTTTCCTGATCCTCCCCTCCCCCACCCTCTCCACCCTACCCTCTGAAAAGGCCCCAGTTGCG 100477

Qy   3135 CATCATTCCACACTTCACATCTATGTTGTACACATTATTAGTGTCCTTACTTATTAAGTGATA 3194
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 100476 TGTTGTTCCCCTTTATGTGTCCANTGTGTTCTCATCATTTACCATTTATAAATGAGA 100417

Qy   3195 ACATGCAATATTTGTCTTTCGTTGTCTGTGTCTGGTCTTGTTGTTTTACTTATGATAAATGCCCCAGTT 3254
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 100416 ACAGTAGTTATTTCTTTTTTGCATTTGTTTGCTAAGGATAAATGGGCCTCCAGCA 100357

Qy   3255 CTATCTAGGCTGTGCAAAAGGCATGAATTCATTCTTTTTTATGGGCTATGTTCTTTCCA 3314
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 100356 CCTTTCANGTTTGCTGCAAGAGGACATGATCCNATTCTTTTTTATGGTTGGCANTAGTATTCCA 100297

Qy   3315 ATTTAGATAAAGAACACTGCACTTGCTCTTACTTCTATTG 3356
            ||||||||||||||||||||||||||||||||||||||||
Db 100296 TGGTGTATATGTACCACATTATCTTTATCCAGTCTATCATTG 100255

RESULT 12
US-08-321-478-2/c
Sequence 2, Application US/08321478
Patent No. 5527677
GENERAL INFORMATION:
APPLICANT: DEGUCHI, Takeo
APPLICANT: KINOSHITA, Moritoshi
APPLICANT: KATSURAGI, Kiyonori
APPLICANT: SHIN, Sadahito
TITLE OF INVENTION: HUMAN ARYLAMINE N-ACETYLTRANSFERASE
TITLE OF INVENTION: GENES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/321,478
FILING DATE: 11-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/038,667
FILING DATE: 23-MAR-1993
APPLICATION NUMBER: JP 64669/1992
FILING DATE: 23-MAR-1992
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 649103
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6464 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 723..1595
FEATURE:
NAME/KEY: exon
LOCATION: 717..1936

FEATURE:
  NAME/KEY: polyA_signal
  LOCATION: 1794..1799
FEATURE:
  NAME/KEY: polyA_signal
  LOCATION: 1800..1805
US-08-321-478-2

Query Match   1.5%;   Score 117.4; DB 1; Length 6464;
Best Local Similarity   63.1%; Pred. No. 1.9e-21;
Matches 181; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```
Qy 3070 CCCTTAAGTAATTTCTCACCATCGGCTGACTTCTTGCCCCCTCCTTCTGAGGCTCCA 3129
Db 5569 CGCATTATTTATTTTTCTGCTTCTCTCTCCTCCTACTCTTCACCCCAGTAGGCCCCA 5510

Qy 3130 TTGTCCATCATTCCACACTCTACACTTATGTGTACACATTATTTAGCTCCTACTTATAAG 3189
Db 5509 GTGTGAGTTGTGTCCCTCCATGTGTCCAATGTGTTCTCATCATTTAGCTTACAAG 5450

Qy 3190 TGATAACATGCAAATAATTTGTCTTCTGTGTCTGTGTTTTACTTATGATAATGCCCC 3249
Db 5449 TAAGAACATGTGGGTATTTGGTTTTCTTGTTTTGCTAAGGATAATAGGCCTC 5390

Qy 3250 CAGTTCTATCTAGGCGTGCTCCAAAAGGCCATGAATTCAGTAATCTTTTAGGGCTATGTTCTT 3309
Db 5389 CAGCTTCATCCATGTTTCTGCAAAGGACATGAATCTCATTCATTCTTTCTTTATAGCTGCATAGTA 5330

Qy 3310 TCCCAATTTAGATAAAGAAGAACACTCGCACTTGCTCTTACTTCTATTTTG 3356
Db 5329 TCCCATGGTGTATATGTACCACACAATTTCTTTATCCAGTGTATCATTG 5283
```

RESULT 13
US-08-321-478-4/c
Sequence 4, Application US/0831478
Patent No. 5527677
GENERAL INFORMATION:
  APPLICANT: DEGUCHI, Takeo
  APPLICANT: KINOSHITA, Moritoshi
  APPLICANT: KATSURAGI, Kiyonori
  APPLICANT: SHIN, Sadahito
  TITLE OF INVENTION: HUMAN ARYLAMINE N-ACETYLTRANSFERASE
  TITLE OF INVENTION: GENES
  NUMBER OF SEQUENCES: 13
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
    STREET: 2100 Pennsylvania Avenue, N.W.
    CITY: Washington
    STATE: D.C.
    COUNTRY: United States
    ZIP: 20037-3202
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/321,478
    FILING DATE: 11-OCT-1994
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US/08/038,667
    FILING DATE: 23-MAR-1993
    APPLICATION NUMBER: JP 64669/1992
    FILING DATE: 23-MAR-1992
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (202) 293-7060
    TELEFAX: (202) 293-7860
    TELEX: 6491103
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
  LENGTH: 6464 base pairs

  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
  NAME/KEY: CDS
  LOCATION: 723..1595
FEATURE:
  NAME/KEY: exon
  LOCATION: 717..1936
FEATURE:
  NAME/KEY: polyA_signal
  LOCATION: 1794..1799
FEATURE:
  NAME/KEY: polyA_signal
  LOCATION: 1800..1805
US-08-321-478-4

Query Match   1.5%;   Score 117.4; DB 1; Length 6464;
Best Local Similarity   63.1%; Pred. No. 1.9e-21;
Matches 181; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```
Qy 3070 CCCTTAAGTAATTTCTCACCATCGGCTGACTTCTTGCCCCCTCCTTCTGAGGCTCCA 3129
Db 5569 CGCATTATTTATTTTTCTGCTTCTCTCTCCTCCTACTCTTCACCCCAGTAGGCCCCA 5510

Qy 3130 TTGTCCATCATTCCACACTCTACACTTATGTGTACACATTATTTAGCTCCTACTTATAAG 3189
Db 5509 GTGTGAGTTGTGTCCCTCCATGTGTCCAATGTGTTCTCATCATTTAGCTTACAAG 5450

Qy 3190 TGATAACATGCAAATAATTTGTCTTCTGTGTCTGTGTTTTACTTATGATAATGCCCC 3249
Db 5449 TAAGAACATGTGGGTATTTGGTTTTCTTGTTTTGCTAAGGATAATAGGCCTC 5390

Qy 3250 CAGTTCTATCTAGGCGTGCTCCAAAAGGCCATGAATTCAGTAATCTTTTAGGGCTATGTTCTT 3309
Db 5389 CAGCTTCATCCATGTTTCTGCAAAGGACATGAATCTCATTCATTCTTTCTTTATAGCTGCATAGTA 5330

Qy 3310 TCCCAATTTAGATAAAGAAGAACACTCGCACTTGCTCTTACTTCTATTTTG 3356
Db 5329 TCCCATGGTGTATATGTACCACACAATTTCTTTATCCAGTGTATCATTG 5283
```

RESULT 14
US-09-822-862-3/c
Sequence 3, Application US/09822862
Patent No. 6468774
GENERAL INFORMATION:
  APPLICANT: GAN, Weiniu et al
  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
  TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
  TITLE OF INVENTION: THEREOF
  FILE REFERENCE: CL001218
  CURRENT APPLICATION NUMBER: US/09/822,862
  CURRENT FILING DATE: 2001-04-02
  NUMBER OF SEQ ID NOS: 4
  SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
  LENGTH: 12385
  TYPE: DNA
  ORGANISM: Human
US-09-822-862-3

Query Match   1.5%;   Score 117; DB 4; Length 12385;
Best Local Similarity   66.7%; Pred. No. 3.9e-21;
Matches 214; Conservative 0; Mismatches 100; Indels 7; Gaps 3;

```
Qy 1170 CAGTAGGGGGTGAAAACGTGTGATACATTTTCTTCCATCATAAGGGTCGCAACCAAAAC 1229
Db 1467 CTGTAGGGACGAGAGAAGGTGGGGTCGGGGTGCCTTTCCTCCCCATCATAAGGATCAACAGCTGACAT 1408

Qy 1230 TCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAAATTTATTTAATCAAAGTT 1289
```

```
Db   1407 GCCTGTAGAAA----AGGTTAGCAAGAGAGAAAAGCATAAC-GATTTATTTAATTAAAGTT 1353
Qy   1290 TTACATGACATGGGAGTGTTTCAGAAATTGAAGACCCAGGGGGAAACTGTCTGTT 1349
Db   1352 TTCTGTGACATGAGAGTCTTCAGAAATGAAGACCCAAAGAACTCAAGAAAACTCTATTTT 1293
Qy   1350 TTTTTTGCTGAGGTTCGATGAAGAAATGGATAGGCATGTAGCCATTGTAGGATTAGACAAAGG 1409
Db   1292 TAGGCTGATTCTATGAAGACCATGGAAAATGC--GCAGAAATGTGGAGTGGGGAAAAGG 1235
Qy   1410 ATATGACTTAGTGTGGTGAAAGGACTCAGGGGGGAAAACACACAGCAAGGCCTGTCTTATTCAGATTC 1469
Db   1234 GAATGGTCTAAATAGTAATAAAACTGAGTGGGAAACCCAGCAAGGCTTGTCTCGTTCGGTTATTC 1175
Qy   1470 TTCTTGATCTCTCTCTCTA 1490
Db   1174 TTCTTGGCCTTATCTGTGGCCA 1154
```

```
RESULT 15
US-08-321-478-6/c
; Sequence 6, Application US/08321478
; Patent No. 5527677
GENERAL INFORMATION:
;   APPLICANT: DEGUCHI, Takeo
;   APPLICANT: KINOSHITA, Moritoshi
;   APPLICANT: KATSURAGI, Kiyonori
;   APPLICANT: SHIN, Sadahito
;   TITLE OF INVENTION: HUMAN ARYLAMINE N-ACETYLTRANSFERASE
;   TITLE OF INVENTION: GENES
;   NUMBER OF SEQUENCES:  13
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
;     STREET: 2100 Pennsylvania Avenue, N.W.
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: United States
;     ZIP: 20037-3202
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/321,478
;     FILING DATE: 11-OCT-1994
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/038,667
;     FILING DATE: 23-MAR-1993
;     APPLICATION NUMBER: JP 64669/1992
;     FILING DATE: 23-MAR-1992
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (202) 293-7060
;     TELEFAX: (202) 293-7860
;     TELEX: 649103
; INFORMATION FOR SEQ ID NO:  6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 6464 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 723..1595
;   FEATURE:
;     NAME/KEY: exon
;     LOCATION: 717..1936
;   FEATURE:
;     NAME/KEY: polyA_signal
;     LOCATION: 1794..1799
;   FEATURE:
```

```
;   FEATURE:
;     NAME/KEY: polyA_signal
;     LOCATION: 1800..1805
US-08-321-478-6

Query Match           1.4%;  Score 115.8;  DB 1;  Length 6464;
Best Local Similarity  62.7%;  Pred. No. 5.3e-21;
Matches  180;  Conservative   0;  Mismatches 107;  Indels   0;  Gaps   0;

Qy   3070 CCCTTAAGTAATTTCTCACCATCCGACTTCTTGCCCCTCATCCTTCTGAGGCTCCA 3129
          ||||| |||| ||| | |||||||||  || | |||  || |  | |||| |||||
Db   5569 CACATTATTTATTTTTCTGTCTTCTCTCTCTCTCTCTTCACCCAGTAGGGCCCA 5510

Qy   3130 TTGTCCATCATTCCACACTCTACATCTATGTGTTACACATTATTTAGCTCCTACTTATAAG 3189
          || ||||||||||||| || |||| |||||||| |||||||| ||||||| ||  |||||
Db   5509 GTGTGGAGTTGTTCCCCTCTTATGGTCCAAGTGTTCCATCATTTAGGTCCACTTACAAG 5450

Qy   3190 TGATAACAATGCAATAATTTGTCTTTCTGTGTCTTGTCTTGTTTTTACTTATGATAAATGGGCCC 3249
          |||| ||||| |||| |||| | ||||||| ||| ||| ||| | || | |||| ||||||||
Db   5449 TAAGAACACTGTGGTTATTTGGTTTTTCTGCATTGGTTTTGCTTAGGGATAAATGGCCCTC 5390

Qy   3250 CAGTGTTCTATCTAGGCTGCTGCAAAAGGCCATGAATTTCAATCTTTTTTTATGGCTAATGTTCTT 3309
          |||| || |||||  | ||| | ||||||||||||| || ||| ||||||| ||||| || ||
Db   5389 CAGCTTCATCCAGTTTCTTGCAAAGGACAAGGACATGGATCTCATTCTTTTCTTAPAGCTGCATAGTA 5330

Qy   3310 TCCCAATTTAGATAAAGAGACACTCGCACTTGCTCTTACTTCTTATTG 3356
          |||||| |||| ||| ||| ||| |||| |||| ||| ||| |||||
Db   5329 TTCCATGGTGTATGTGGTACCACACATTTCTTTACCAGTGTATCATTG 5283

Search completed: September 11, 2004, 20:14:21
Job time : 357 secs
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: September 11, 2004, 16:46:54 ; Search time 2304 Seconds

    GenCore version 5.1.6

Title: US-09-939-209A-3_COPY_4110_12150
Perfect score: 8041
Sequence: 1 caaattgaagtcctgggcac.........ctaggactcaggtggatcc 8041

Scoring table: IDENTITY_NUC
    Gapop 10.0 , Gapext 1.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
    Maximum Match 100%
    Listing first 45 summaries

Database :   Published Applications NA:*
  1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
  2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
  3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
  4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
  5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
  6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
  7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
  8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
  9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
  10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
  11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
  12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
  13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
  14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
  15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
  16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
  17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
  18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
  19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

(without alignments)
17560.102 Million cell updates/sec

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB ID	Description
1	100.0	20300	10	US-09-939-209A-3	Sequence 3, Appli
2	600.8	7.5	666	13	US-10-027-632-68745	Sequence 68745, A
3	600.8	7.5	666	16	US-10-027-632-68745	Sequence 68745, A
4	598.2	7.4	615	13	US-10-027-632-68208	Sequence 68208, A
5	598.2	7.4	615	13	US-10-027-632-68209	Sequence 68209, A
6	598.2	7.4	615	13	US-10-027-632-68210	Sequence 68210, A
7	598.2	7.4	615	13	US-10-027-632-68211	Sequence 68211, A
8	598.2	7.4	615	13	US-10-027-632-68742	Sequence 68742, A
9	598.2	7.4	615	13	US-10-027-632-68743	Sequence 68743, A
10	598.2	7.4	615	13	US-10-027-632-68744	Sequence 68744, A
11	598.2	7.4	615	13	US-10-027-632-70068	Sequence 70068, A
12	598.2	7.4	615	13	US-10-027-632-70069	Sequence 70069, A
13	598.2	7.4	615	13	US-10-027-632-70070	Sequence 70070, A
14	598.2	7.4	615	13	US-10-027-632-70071	Sequence 70071, A
15	598.2	7.4	615	13	US-10-027-632-295009	Sequence 295009,
16	598.2	7.4	615	13	US-10-027-632-295010	Sequence 295010,
17	598.2	7.4	615	13	US-10-027-632-311585	Sequence 311585,
18	598.2	7.4	615	13	US-10-027-632-311586	Sequence 311586,
19	598.2	7.4	615	16	US-10-027-632-68208	Sequence 68208, A
20	598.2	7.4	615	16	US-10-027-632-68209	Sequence 68209, A
21	598.2	7.4	615	16	US-10-027-632-68210	Sequence 68210, A
22	598.2	7.4	615	16	US-10-027-632-68211	Sequence 68211, A
23	598.2	7.4	615	16	US-10-027-632-68742	Sequence 68742, A
24	598.2	7.4	615	16	US-10-027-632-68743	Sequence 68743, A
25	598.2	7.4	615	16	US-10-027-632-68744	Sequence 68744, A
26	598.2	7.4	615	16	US-10-027-632-70068	Sequence 70068, A
27	598.2	7.4	615	16	US-10-027-632-70069	Sequence 70069, A
28	598.2	7.4	615	16	US-10-027-632-70070	Sequence 70070, A
29	598.2	7.4	615	16	US-10-027-632-70071	Sequence 70071, A
30	598.2	7.4	615	16	US-10-027-632-295009	Sequence 295009,
31	598.2	7.4	615	16	US-10-027-632-295010	Sequence 295010,
32	598.2	7.4	615	16	US-10-027-632-311585	Sequence 311585,
33	598.2	7.4	615	16	US-10-027-632-311586	Sequence 311586,
34	277.6	3.5	2752	15	US-10-198-846-13843	Sequence 13843, A
35	276	3.4	1531	15	US-10-198-846-9863	Sequence 9863, Ap
36	274.4	3.4	3165	14	US-10-044-099-113	Sequence 113, App
37	274.4	3.4	3167	10	US-10-027-632-266349	Sequence 6, Appli
38	253.6	3.2	1151	16	US-10-027-632-266349	Sequence 266349,
39	253.6	3.2	1151	16	US-09-918-624B-6	Sequence 266349,
40	253.2	3.1	1151	13	US-10-027-632-266349	Sequence 266350,
41	253.2	3.1	1151	13	US-10-027-632-266350	Sequence 266350,
42	233.8	2.9	576	16	US-10-027-632-288758	Sequence 288758,
43	233.8	2.9	576	16	US-10-027-632-288758	Sequence 288758,
44	233.4	2.9	576	13	US-10-027-632-288755	Sequence 288755,
45	233.4	2.9	576	13	US-10-027-632-288756	Sequence 288756,

ALIGNMENTS

RESULT 1
US-09-939-209A-3

Sequence 3, Application US/09939209A
Publication No. US20030113721A1
GENERAL INFORMATION:
  APPLICANT: LEVITT, PAT R
  APPLICANT: MIRNICS, KAROLY
  APPLICANT: KODAVALI, MENKATA CHOWDARI
  APPLICANT: NIMGAONKAR, VISHWAJIT L.
  TITLE OF INVENTION: METHODS AND SYSTEMS FOR FACILITATING THE DIAGNOSIS AND TREATMENT
  TITLE OF INVENTION: SCHIZOPHRENIA
  FILE REFERENCE: 00-539-US
  CURRENT APPLICATION NUMBER: US/09/939,209A
  CURRENT FILING DATE: 2001-08-24
  PRIOR APPLICATION NUMBER: 60/228,021
  PRIOR FILING DATE: 2000-08-24
  NUMBER OF SEQ ID NOS: 8
  SOFTWARE: Patentin version 3.1

SEQ ID NO 3
  LENGTH: 20300
  TYPE: DNA
  ORGANISM: Artificial Sequence
  FEATURE:
  NAME/KEY: misc feature
  OTHER INFORMATION: A genomic sequence containing RGS4 nucleic acid sequence and seq
  OTHER INFORMATION: ences upstream and downstream to the RGS4 nucleic acid sequence
US-09-939-209A-3

Query Match   100.0%;   Score 8041;   DB 10;   Length 20300;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 8041;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 CAAATTGAAGTCCTGGGCACGTTACAAAACTAGAGAGTCATAAAGTTTACCCTAAATTTACC 60
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4110 CAAATTGAAGTCCTGGGCACGTTACAAAACTAGAGAGTCATAAAGTTTACCCTAAATTTACC 4169

Qy   61 AAGATTTCCTAGAGGATCATAATTGGAATCACCAGATCTCGCCTCTCTGTAAAGTTCAAGCA 120

(This page consists of a two-column DNA sequence alignment between Qy and Db sequences, with position numbers flanking each 60-base row. The left column spans positions 2281–7469 and the right column spans positions 3361–4500, each row labeled "Qy" or "Db".)

```
Db  8550 TATATTTAACTCATGACAAATCAGCTTCTCTTATTCTAAGTCAAGACATAAAAACTTTT 8609
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4501 ATTGTGGAATTCAGCATCAGTAAATCTTTTGGGTACTCACTTAGTTCCTGAAATCA 4560

Db  8610 ATTGTGGAATTCAGCCATCAGTAAATCTTTTGGGTACTCACTTAGTTCCTGAAATCA 8669
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4561 ATCTATTTGAGTGATCACTCTTTTAGGTGTGCCCAGGTAAACAAAGGCCATGGTCTTTC 4620

Db  8670 ATCTATTTGAGTGATCACTCTTTTAGGTGTGCCCAGGTAAACAAAGGCCATGGTCTTTC 8729
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4621 TTTGAGTGACCTTCTTTCCCTTTTAATTAGTCTGACCTCTTTTAATGTCAGTTCTGACTGA 4680

Db  8730 TTTGAGTGACCTTCTTTCCCTTTTAATTAGTCTGACCTCTTTTAATGTCAGTTCTGACTGA 8789
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4681 TTCATTTCCCTGGTCATCTTCCTTGGGTCTGAGGGCCCTTCACCTCTTTTAAGTGTCACTTC 4740

Db  8790 TTCATTTCCCTGGTCATCTTCCTTGGGTCTGAGGGCCCTTCACCTCTTTTAAGTGTCACTTC 8849
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4741 AGTTCCTTCCACACCACCATCAAGGATGGCTGTGTCAACAATCAATTGTTCTAGTTATAAT 4800

Db  8850 AGTTCCTTCCACACCACCATCAAGGATGGCTGTGTCAACAATCAATTGTTCTAGTTATAAT 8909
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4801 TCAAGGAAAAGTTGCCCAGTAGCTAATCCAATAAAATGCCCTTCTTAGGGGCGGCTAGAGAC 4860

Db  8910 TCAAGGAAAAGTTGCCCAGTAGCTAATCCAATAAAATGCCCTTCTTAGGGGCGGCTAGAGAC 8969
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4861 TTTTTCCTATAATTTAAATGCATCTTTCTGTAGATTATATGGTCCCTCCACCACTTTACATTT 4920

Db  8970 TTTTTCCTATAATTTAAATGCATCTTTCTGTAGATTATATGGTCCCTCCACCACTTTACATTT 9029
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4921 GTCTGCTGTCTCCCTTGCTCTGCTAGTCATGGAACGTGTTGGGTAGTGTGGGGCAGTGTGGGA 4980

Db  9030 GTCTGCTGTCTCCTTGCTCTGCTAGTCATGGAACGTGTTGGGTAGTGTGGGGCAGTGTGGGA 9089
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  4981 ATATTTTTGGTATTTTTGCATCTCACTGGAACCCAACTATTTTTCATCCTTCCACCTAAA 5040

Db  9090 TGTTCAAGGGCACGTATTGGGGTAGGGCCCACATATGGGGCATTCATCCTTCCACCTAAA 9149
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5041 ATATTTTTGGTATTTTTGCATCTCACTGGAACCCAACTATTTTTCATCCTTCCACCTAAA 5100

Db  9150 ATATTTTTGGTATTTTTGCATCTCACTGGAACCCAACTATTTTTCATCCTTCCACCTAAA 9209
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5101 CTATTTGATGCCTCTGTTTCTTATATATAAAGTATAGTAGCCTATGATCAGGA 5160

Db  9210 CTATTTGATGCCTCTGTTTCTTATATATAAAGTATAGTAGCCTATGATCAGGA 9269
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5161 ACCTATCTGCTTTCTAAATGAAAGCTGTTTTGGTCAGCAATTAAATTCCCTTCTT 5220

Db  9270 ACCTATCTGCTTTCTAAATGAAAGCTGTTTTGGTCAGCAATTAAATTCCCTTCTT 9329
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5221 CCACTTATAGCTTTCCTCTGTAACTCTGGTGTAGGTATATTGGTTTATGGCCTATAAGATGT 5280

Db  9330 CCACTTATAGCTTTCCTCTGTAACTCTGGTGTAGGTATATTGGTTTATGGCCTATAAGATGT 9389
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5281 GAAAACACCTGAATGATTCTGTCCATGCAGGCATTCAGTTCATGATGATTGTATGTAAAAAG 5340

Db  9390 GAAACACCTGAATGATTCTGTCCATGCAGGCATTCAGTTCATGATGATTGTATGTAAAAAG 9449
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5341 ATACTGATTGTCTAGGTGTTCAGAAAACACCTATAGGGGCTTAAATATTCTTACAATCAGTTT 5400

Db  9450 ATACTGATTGTCTAGGTGTTCAGAAAACACCTATAGGGGCTTAAATATTCTTACAATCAGTTT 9509
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5401 GAAGGCTGGTGATAGCAAAGACTACATATATTTTTCTGCCTGCTCTCTCTTTTCTCT 5460

Db  9510 GAAGGCTGGTGATAGCAAAGACTACATATATTTTTCTGCCTGCTCTCTCTTTTCTCT 9569
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5461 CTACACTCTCTTCTTTATCTTTGAAATACAGTTTGGACAGTTACATACAATAAGA 5520

Db  9570 CTACACTCTCTCTTTATCTTTGAAATATACAGTTCTGTGCAAATTAGAAACAGTGTGGTCAG 9629
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5521 CATAAAACCCATTGATATAAGAATTGCTGTGTATATATTGGCTCATCTACTCCTCCTTGG 5580

Db  9630 CATAAAACCCATTGATATAAGAATTGCTGTGTATATATTGGCTCATCTACTCCTCCTTGG 9689
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5581 TCCTCGGAGCTCGCCCGGTTTAGACTTTTTTACAGGATCAAGGACATGTGAGGCAGAAACTCTCA 5640

Db  9690 TCCTCGGAGCTCGCCGGTTTAGACTTTTTTACAGGATCAAGGACATGTGAGGCAGAAACTCTCA 9749
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5641 GTGCTTAGGCTGGAATTCTGTGTTGTTACCAAGAATTTCTAGAAAAGTATTCCTCAGTCAGGTTG 5700

Db  9750 GTGCTAGGCTGGAATTCTGTGTTGTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTG 9809
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5701 ATTACAGATATAGCAAATCTTATTTTTCCTAGGGTAGTTCTGTATGGTCCTTATA 5760

Db  9810 ATTACAGATATAGCAAAATCTATTTTTCCTAGGGTAGTTCTGTATGGTCCGGGGCTTATA 9869
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5761 ACTGTTCTGTCATCCAGCTATTTCTCTCCACCTTCTGTTTGTTGCTAACAACCAAGGCAAACT 5820

Db  9870 ACTGTTCTGTCATCCAGCTATTTCTCTCCACCTTCTGTTTGTTGCTAACAACCAAGGCAAACT 9929
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5821 TCCCCAAAATCACTGCGTGGAGGACGATGAATCCTGCCAGCTCCCTTTGGAAATCGTGAGGA 5880

Db  9930 TCCGCAAATCACTGCGTGGAGGACGATGAATCCTGCCAGCTCCCTTTGGAAATCGTGAGGA 9989
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  5881 TCAGATCTTGGACCATGTAAATGATGGCTTCTAATCCAAAAGAGGAAAAGGCATTGGGA 5940

Db  9990 TCAGATCTTGGACCATGTATAAATGATGGCTTCTAATCCAAAAGAGGAAAAGGCATTGGGA 10049

Db  10050 GTCAGCTCCTTAAGTAAGCTCCAGAATTCCTGCGTCGTGTACTTTTCCTTCCAGGAAGCAACTT 10109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6001 GTCAGCTCCTTAAGTAAGCTCCAGAATTCCTGCGTCGTGTACTTTTCCTTCCAGGAAGCAACTT 6060

Db  10110 CCTTGATATATTTTTTTTTTACAGGCATATGAATGAATAAAACTATATTTGTTGCATTGTACAC 10169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6061 TTTTTTTTTCTTTTTCTTTTCTAGAAAATTCTAAACCTCTGACATTGGTTGGGAGACATTGATT 6120

Db  10170 TTTTTTTTTCTTTTTCTTTTCTAGAAATTCTAAACCTCTGACATTGGTTGGGAGACATTGATT 10229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6121 TTTCCCATATCCCTACTTTTCAGAAGGAGATTTTCTCTCGTCGTTCACTTAACATTGCTTGAT 6180

Db  10230 TTTCCCATATCCCTACTTTTCAGAAGGAGATTTTCTCTCGTCGTTCACTTAACATTGCTTGAT 10289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6181 GCGGTCAGGTCTTTTCTTTCCCATCTCTTCAGGGGTGGGAGGGCAGAGGGGGAGACAGAGGA 6240

Db  10290 GCGGTCAGGTCTTTTCTTTCCCATCTCTTCAGGGGTGGGAGGGCAGAGGGGGAGACAGAGGA 10349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6241 GCTGGTACTTGCAGAGGCGGTCGTCTGATTGGCTGGACGGTCGTAGCGTATAAAGAG 6300

Db  10350 GCTGGTACTTGCAGAGGCGGTCGTCTGATTGGCTGGACGGTCGTAGCGTATAAAGAG 10409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6301 ACCCCTACAGGCCTTAGCAGGAGAACGCTCAGAGGAGATTCTGACAAATATCTTTACCGGAA 6360

Db  10410 ACCCCTACAGGCCTTAGCAGGAGAACGCTCAGAGGAGATTCTGACAAATATCTTTACCGGAA 10469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6361 GAGGCAAAGTACGCTCAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTTCCTGCTT 6420

Db  10470 GAGGCAAAGTACGCTCAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTTCCTGCTT 10529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6421 GCGGAATTCCAAGCTGTTAAATAAGAGTGCAAAGGGCTTGCAGGGTCGGGCTTTGC 6480

Db  10530 GCGGAATTCCAAGCTGTTAAATAAGAGTGCAAAGGGCTTGCAGGGTCGGGCTTTGC 10589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6481 TTGAGGAGGTAAGCTAAGGTTGCTTTCAGCCATTAACCATATTAAAACTTTTGGCTAGAACTTTCTC 6540

Db  10590 TTGAGGAGGTAAGCTAAGGTTGCTTTCAGCCATTAACCATATTAAAACTTTTGGCTAGACTTTCTC 10649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6541 AGTTATTTACATGTTGTACTTACTAACCTAGTTCTGTTGCAATTAGAAACAGTGTGGTCAG 6600

Db  10650 AGTTATTTACATGTTGTACTTACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAG 10709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   6601 GAGAGCAGCAGACTTTCTAACTTTCCTCCCAAGACTAGCTAGATATATTTGTGACTTAAGACATGT 6660

Db  10710 GAGAGCAGCAGACTTTCTAACTTTCCTCCCAAGACTAGCTAGATATATTTGTGACTTAAGACATGT 10769
```

```
Qy   6661  GCTCCCCAAATTTCAGCCCCTTATCTGTGTTGTTTGTGTCGACCTCAGTTTGAGAACTGTTC  6720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10770  GCTCCCCAAATTTCAGCCCCTTATCTGTGTTGTTTGTGTCGACCTCAGTTTGAGAACTGTTC 10829

Qy   6721  TATTCCTTTAAGCCAGGTCTAAGAAAGCTAGTAGTTTAATTAATAGAAGAGAGAGGTTGAG  6780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10830  TATTCCTTTAAGCCAGGTCTAAGAAAGCTAGTAGTTTAATTAATAGAAGAGAGAGGTTGAG 10889

Qy   6781  GCTAGTACAGTGATCTGTAATATCTTCCATCTGTGTACTGCTATTTGAGCATCCC  6840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10890  GCTAGTACAGTGATCTGTAATATCTTCCATCTGTGTACTGCTATTTGAGCATCCC 10949

Qy   6841  TGGAGTACATAGAAGCCCTGGCTCTGGGCGTTCTGATTGTATGCTCACAACTTGTTCAGGA  6900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10950  TGGAGTACATAGAAGCCCTGGCTCTGGGCGTTCTGATTGTATGCTCACAACTTGTTCAGGA 11009

Qy   6901  AAGGTACCCCAGAATGAGGTTTGGCTCATCCATCAGAAGAGGCACTATGGCTTTTCGGT  6960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11010  AAGGTACCCCAGAATGAGGTTTGGCTCATCCATCAGAAGAGGCACTATGGCTTTTCGGT 11069

Qy   6961  TGGTGCCAGTAACTTTCACTCTGTCCTAGTTCTTATAAGCAAAAGTTACAATGAGATACAGT  7020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11070  TGGTGCCAGTAACTTTCACTCTGTCCTAGTTCTTATAAGCAAAAGTTACAATGAGATACAGT 11129

Qy   7021  TTTAAAGCCAGAGATCTTCCTTATCTCTGCCCATCTCAGTTGTTCTTGAAGTGTGTCATAT  7080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11130  TTTAAAGCCAGAGATCTTCCTTATCTCTGCCCATCTCAGTTGTTCTTGAAGTGTGTCATAT 11189

Qy   7081  GAGTTTGGGTTGAGAAATATTGATCATTACAAAAATCAGTTAATAGTTTTGTAGAAGATCTCA  7140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11190  GAGTTTGGGTTGAGAAATATTGATCATTACAAAATCAGTTAATAGTTTTGTAGAAGATCTCA 11249

Qy   7141  TCTTAAAGACATTGTTTTGTTAATATTACTCCCCTTGATTTTTTAAAAGACCCTTACAGACA  7200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11250  TCTTAAAGACATTGTTTTGTTAATATTACTCCCCTTGATTTTTTAAAAGACCCTTACAGACA 11309

Qy   7201  TACAGCTATTCATTGTTGTTTTGGTTTGTTCAAAAAGGTATAAAGAAAAGCATTCAGAGA  7260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11310  TACAGCTATTCATTGTTGTTTTGGTTTGTTCAAAAAGGTATAAAGAAAAGCATTCAGAGA 11369

Qy   7261  AAGATCATATATTAGCCAGTTGAAAATTAAACAAAATGAGTGCATATTACATTACTTA  7320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11370  AAGATCATATATTAGCCAGTTGAAAATTAAACAAAATGAGTGCATATTACATTACTTA 11429

Qy   7321  ATCTTGCAGTCAGTAAAGGTAAACCTAACCTGCCTTCCTTTTATCGCAC  7380
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db  11430  ATCTTGCAGTCAGTAAAGGTAAACCTAACCTGCCTTCCTTTTATCGGCAC 11489

Qy   7381  TGCAAATAGAAAATTACCACAAAAATTTTATTTTGGAAATAAATCTCGAAAAACACAATAATTTTTT  7440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11490  TGCAAATAGAAAATTACCACAAAAATTTTATTTTGGAAATAAATCTCGAAAAACACAATAATTTTTT 11549

Qy   7441  ATGTACTAATTAAAACATTTTACTTTCCAAATATTCTGTCATTCAGGAGTAGAAGTAATCG  7500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11550  ATGTACTAATTAAAACATTTTACTTTCCAAATATTCTGTCATTCAGGAGTAGAAGTAATCG 11609

Qy   7501  ATGGCTTCTTAAAATGAACAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCT  7560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11610  ATGGCTTCTTTAAAATGAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCT 11669

Qy   7561  CTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAAATTACTTCAAGCTTCACTTTA  7620
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11670  CTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAAATTACTTCAAGCTTCACTTTA 11729

Qy   7621  CTTGCATATGAAATGAACCGTACAAAAATTACAGCAGTGTTCAGGAGAAAGTTATGCTCTGG  7680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11730  CTTGCATATGAAATGAACCGTACAAAAATTACAGCAGTGTTCAGGAGAAAGTTATGCTCTGG 11789

Qy   7681  TAAAATATTTTGCAAAAACAGATAAAAGATAAATAACTTAGAGAGCTCTGTCCCTCAAAGAGTTAAGC  7740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11790  TAAAATATTTTGCAAAAACAGATAAAAGATAATAACTTAGAGAGCTCTGTCCCTCAAAGAGTTAAGC 11849
```

```
Qy   7741  AGCTAATCTAAGGAGGTAAAACTTATGTCAGCAGAGGCGAGGAGGATGAAACGCTCTTCCGTTTCTCCTCCGCCT  7800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11850  AGCTAATCTAAGGAGGTAAAACTTATGTCAGCAGAGGCGAGGAGGATGAAACGCTCTTCCGTTTCTCCTCCGCCT 11909

Qy   7801  CAATAAATTGCAAATCATCCTAGTCCAAACTCTTATTTACCAGTCGCCTTGAGGTCTCCAGAGG  7860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11910  CAATAAATTGCAAATCATCCTAGTCCAAACTCTTATTTACCAGTCGCCTTGAGGTCTCCAGAGG 11969

Qy   7861  AGCCAATTGCCTTCTTCCAAGGTCACATAGGTGGGTCGGGGGGTGAGGTTAGGGACCAAAATCTTAGAATTC  7920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11970  AGCCAATTGCCTTCTTCCAAGGTCACATAGGTGGGTCGGGGGGTGAGGTTAGGGACCAAAATCTTAGAATTC 12029

Qy   7921  CTGACTTCCAGTAACTTCTGAAGTCGATCATTTTGTTTTTTGTTTTTTATTGGTTTTTTTATTATTGGTTTTTATTATATAAGAA  7980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12030  CTGACTTCCAGTAACTTCTGAAGTCGATCATTTTGTTTTTTGTTTTTTATTGGTTTTTTATTATATAAGAA 12089

Qy   7981  TACTTCCTAAGCACAACTTACCCCCTGCATTGATTAATAACTCTTAGGATCTTCAGGTGGATC  8040
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12090  TACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTTAGGATCTTCAGGTGGATC 12149

Qy   8041  C  8041
           |
Db  12150  C 12150
```

RESULT 2
US-10-027-632-68745
Sequence 68745, Application US/10027632
Publication No. US20020198371A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108227.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US/10/027,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 68745
    LENGTH: 666
    TYPE: DNA
    ORGANISM: Human
US-10-027-632-68745

```
Query Match   98.1%;   Score 600.8;   DB 13;   Length 666;
Best Local Similarity   7.5%;
Matches  614;  Conservative   4;  Mismatches   6;  Indels   2;  Gaps   1;

Qy   393  AAGCTTGGTCGCAAAAGAATTGGACACATTTCCCAAAAGTAAGACATAGTGGGAAGTCCCTG  452
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  AAGCTTGGTCGCAAAAGAATTGGACACATTTCCCAAAAGTAAGACATAGTGGGAAGTCCCTG   60

Qy   453  TTTACTTCCTGGTATACAGCACTCCCAGCCCCATAATCTTTGCTTTTTAGTCCTTCCAAAAA  512
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATAACCTTTGCTTTTTAGTCCTTCCAAAAAM  120

Qy   513  TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGAAGAGGA  572
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGAAGAGGA  180
```

RESULT 3
US-10-027-632-68745
Sequence 68745, Application US/10027632
Publication No. US20030240075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 68745
LENGTH: 666
TYPE: DNA
ORGANISM: Human
US-10-027-632-68745
Matches 614; Conservative 4; Mismatches 6; Indels 2; Gaps 1;
Best Local Similarity 98.1%; Pred. No. 3.1e-134;
Query Match 7.5%; Score 600.8; DB 16; Length 666;

Qy 393 AAGCTTGGTGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 AAGCTTGGGTGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

RESULT 4
US-10-027-632-68208
Sequence 68208, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 68208
LENGTH: 615
TYPE: DNA

```
; ORGANISM: Human
US-10-027-632-68208

Query Match      7.4%;  Score 598.2;  DB 13;  Length 615;
Best Local Similarity 98.7%;  Pred. No. 1.2e-133;
Matches 609;  Conservative 4;  Mismatches 2;  Indels 2;  Gaps 1;

Qy  393 AAGCTTGGTGCAAAGAATTGGACACATTCCCAAAAAGTAAGACATACTGGGAAGTCCCTG 452
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AAGCTTGGTGCAAAGAATTGGACACATTCCCAAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy  453 TTTACCTTTCCTGGTATACAGCATCCTCCAGCCCCATAATCTTTGCTTTTTTAGTCCTAAAAA 512
            |||||||||||||||||||||||||||||||||||||||| ||||||||||:||
Db   61 TTTACCTTTCCTGGTATACAGCATCCTCCAGCCCCATAATCTTTGCTTTTTTAGTCCTAAAAM 120

Qy  513 TCAATAACTGAACTCTCATTGAGTGTCTAGGCCATTGTAGTGAATAAACAATAAAAGAAGGAGGGA 572
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 TCAATAACTGAACTCTCATTGAGTGTCTAGGCCATTGTAGTGAATAAACAATAAAAGAAGGAGGGA 180

Qy  573 GGCTTCTGACAACTGAGAGGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGGCTG 632
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GGCTTCTGACAACTGAGAGGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGGCTG 240

Qy  633 AGCCTTGGCCTACATCCTGCCCAAGTGGGAGGATCAGTAGTGCCCAATTTAACATCTGGTAGAA 692
            |||||||||||||||||||||||||||||||||||||||||||||:|||:|||
Db  241 AGCCTTGGCCTACATCCTGCCCAAGTGGGAGGATCAGTAGTGCCCAATTTAAYATCTGGTAGAA 300

Qy  693 CTAAAGAAGGCAAGGCCCTCCCACAAATGACTTATTTCCCTGCATTGAATAACGTCAATCCT 752
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CTAAAGAAGGCAAGGCCCTCCCACAAATGACTTATTTCCCTGCATTGAATAACGTCAATCCT 360

Qy  753 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAAGGTTGGGAAAAATTGAAAATTTACCTA 812
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAAGGTTGGGAAAAATTGAAAATTTACCTA 420

Qy  813 GAGACCACACTAGTTCACATCCTGCTGTGGGCTGAATGTCTGCCCCCCAGTAGGAGAAAC 872
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GAGACCACACTAGTTCACATCCTGCTGTGGGCTGAATGTCTGCCCCCCAGTAGGAGAAAC 480

Qy  873 AGTTCTTCTAAAGCCTATTGTCAACAATAACCTTCCAGATGTTAGCATTTACAAATTTAAG 932
            |||||||||||||||||||||||||||||||||||||:||--|||||||
Db  481 AGTTCTTCTAAAGCCTATTGTCAACAATAACCTTCCAGATGTTAGCATTTT--CATTTAAG 538

Qy  933 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTTCTCTGATATCCAATCTATTCTTTTA 992
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  539 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTTCTCTGATATCCAATCTATTCTTTTA 598

Qy  993 CTCTGCCCTCCCAAGCTT 1009
            |||||||||||||||||
Db  599 CTCTGCCCTCCCAAGCTT 615
```

RESULT 5

```
Query Match      7.4%;  Score 598.2;  DB 13;  Length 615;
Best Local Similarity 98.7%;  Pred. No. 1.2e-133;
Matches 609;  Conservative 4;  Mismatches 2;  Indels 2;  Gaps 1;

Qy  393 AAGCTTGGTGCAAAGAATTGGACACATTCCCAAAAAGTAAGACATACTGGGAAGTCCCTG 452
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AAGCTTGGTGCAAAGAATTGGACACATTCCCAAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy  453 TTTACCTTTCCTGGTATACAGCATCCTCCAGCCCCATAATCTTTGCTTTTTTAGTCCTAAAAA 512
            |||||||||||||||||||||||||||||||||||||||| ||||||||||:||
Db   61 TTTACCTTTCCTGGTATACAGCATCCTCCAGCCCCATAATCTTTGCTTTTTTAGTCCTAAAAM 120

Qy  513 TCAATAACTGAACTCTCATTGAGTGTCTAGGCCATTGTAGTGAATAAACAATAAAAGAAGGAGGGA 572
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 TCAATAACTGAACTCTCATTGAGTGTCTAGGCCATTGTAGTGAATAAACAATAAAAGAAGGAGGGA 180

Qy  573 GGCTTCTGACAACTGAGAGGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGGCTG 632
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GGCTTCTGACAACTGAGAGGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGGCTG 240

Qy  633 AGCCTTGGCCTACATCCTGCCCAAGTGGGAGGATCAGTAGTGCCCAATTTAACATCTGGTAGAA 692
            |||||||||||||||||||||||||||||||||||||||||||||:|||:|||
Db  241 AGCCTTGGCCTACATCCTGCCCAAGTGGGAGGATCAGTAGTGCCCAATTTAAYATCTGGTAGAA 300

Qy  693 CTAAAGAAGGCAAGGCCCTCCCACAAATGACTTATTTCCCTGCATTGAATAACGTCAATCCT 752
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CTAAAGAAGGCAAGGCCCTCCCACAAATGACTTATTTCCCTGCATTGAATAACGTCAATCCT 360

Qy  753 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAAGGTTGGGAAAAATTGAAAATTTACCTA 812
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAAGGTTGGGAAAAATTGAAAATTTACCTA 420

Qy  813 GAGACCACACTAGTTCACATCCTGCTGTGGGCTGAATGTCTGCCCCCCAGTAGGAGAAAC 872
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GAGACCACACTAGTTCACATCCTGCTGTGGGCTGAATGTCTGCCCCCCAGTAGGAGAAAC 480

Qy  873 AGTTCTTCTAAAGCCTATTGTCAACAATAACCTTCCAGATGTTAGCATTTACAAATTTAAG 932
            |||||||||||||||||||||||||||||||||||||:||--|||||||
Db  481 AGTTCTTCTAAAGCCTATTGTCAACAATAACCTTCCAGATGTTAGCATTTT--CATTTAAG 538

Qy  933 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTTCTCTGATATCCAATCTATTCTTTTA 992
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  539 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTTCTCTGATATCCAATCTATTCTTTTA 598

Qy  993 CTCTGCCCTCCCAAGCTT 1009
            |||||||||||||||||
Db  599 CTCTGCCCTCCCAAGCTT 615
```

RESULT 6

; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68210
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68210

Query Match   7.4%; Score 598.2; DB 13; Length 615;
Best Local Similarity   98.7%; Pred. No 1.2e-133;
Matches   609; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

Qy 393 AAGCTTGGTGCAAAGAATTGGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGCTTGGTGCAAAGAATTGGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy 453 TTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTTAGTCTCTAAAA 512
        |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   61 TTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTTAGTCTCTAAAAM 120

Qy 513 TCAATAACTGAAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAGAAGGAGGGA 572
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 TCAATAACTGAAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAGAAGGAGGGA 180

Qy 573 GGCTTCTGACAACTGAGAGGAAAATGTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 632
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 GGCTTCTGACAACTGAGAGGAAAATGTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy 633 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 692
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy 693 CTAAAGAAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752
        ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db 301 CTAAAGAAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy 753 TGAGAAATGTTTTCTTTTTGTTCTCCCTGACCAAAGGTTGGAAAAATTTGAAATTTACCTA 812
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 TGAGAAATGTTTTCTTTTTGTTCTCCCTGACCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy 813 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAAATGTCTGCCCCCAGTAGGAAAC 872
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAAATGTCTGCCCCCAGTAGGAAAC 480

Qy 873 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTTACAATTTAAG 932
        |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db 481 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTT--CATTTAAG 538

Qy 933 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAATCTATTCTTTTA 992
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 539 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAATCTATTCTTTTA 598

Qy 993 CTCTGCCTCCCAAGCTT 1009
        |||||||||||||||||
Db 599 CTCTGCCTCTCAAGCTT 615

RESULT 7
US-10-027-632-68211
; Sequence 68211, Application US/10027632

Publication No. US20020193571A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68211
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68211

Query Match   7.4%; Score 598.2; DB 13; Length 615;
Best Local Similarity   98.7%; Pred. No 1.2e-133;
Matches   609; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

Qy 393 AAGCTTGGTGCAAAGAATTGGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGCTTGGTGCAAAGAATTGGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy 453 TTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTTAGTCTCTAAAA 512
        |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   61 TTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTTAGTCTCTAAAAM 120

Qy 513 TCAATAACTGAAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAGAAGGAGGGA 572
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 TCAATAACTGAAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAGAAGGAGGGA 180

Qy 573 GGCTTCTGACAACTGAGAGGAAAATGTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 632
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 GGCTTCTGACAACTGAGAGGAAAATGTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy 633 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 692
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy 693 CTAAAGAAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752
        ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db 301 CTAAAGAAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy 753 TGAGAAATGTTTTCTTTTTGTTCTCCCTGACCAAAGGTTGGAAAAATTTGAAATTTACCTA 812
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 TGAGAAATGTTTTCTTTTTGTTCTCCCTGACCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy 813 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAAATGTCTGCCCCCAGTAGGAAAC 872
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAAATGTCTGCCCCCAGTAGGAAAC 480

Qy 873 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTTACAATTTAAG 932
        |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db 481 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTT--CATTTAAG 538

Qy 933 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAATCTATTCTTTTA 992
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 539 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAATCTATTCTTTTA 598

```
Qy    993 CTCTGCCTCTCCAAGCTT 1009
          ||||||||||||||||||
Db    599 CTCTGCCTCTCCAAGCTT 615
```

RESULT 8
US-10-027-632-68742
Sequence 68742, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 10827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 68742
LENGTH: 615
TYPE: DNA
ORGANISM: Human
US-10-027-632-68742

```
Query Match             98.7%;  Score 598.2;  DB 13;  Length 615;
Best Local Similarity   98.7%;  Pred. No. 1.2e-133;
Matches  609;  Conservative   4;  Mismatches   2;  Indels   2;  Gaps   1;

Qy  393 AAGCTTGGTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AAGCTTGGTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy  453 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCCATATTCTTTGCTTTTTTAGTCCTAAAAA 512
        |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||:
Db   61 TTTACCTTCCTGGTATACACCATCCTCCAGCCCCCATATTCTTTGCTTTTTTAGTCCTAAAAM 120

Qy  513 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGAGGGA 572
        |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  121 TCAATAACTGAACTCTCATTGAGTCTAGGCCATTGTGAGTAAACAATAAAGAAGGAGGAGGGA 180

Qy  573 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCAGCCTGGGGCTG 632
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCAGCCTGGGGCTG 240

Qy  633 AGCCTTGGCCTACATCCTGCCCCAAGTGGGAGGATCAGTGCCCCCATTTAAACATCTGGTAGA 692
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AGCCTTGGCCTACATCCTGCCCCAAGTGGGAGGATCAGTGCCCCCATTTAAACATCTGGTAGA 300

Qy  693 CTAAAGAACGCAAGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CTAAAGAACGCAAGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy  753 TGAGAAAATGTTTTTCTTTGTTCTCCCTGAGAGTAAGTTGGAAAAATTTGAAATTTACCTA 812
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TGAGAAAATGTTTTTCTTTGTTCTCCCTGAGAGTAAGTTGGAAAAATTTGAAATTTACCTA 420

Qy  813 GAGACCACACAATAGTTCACATCCTGCTGTGTGGCTGAAGTGCTGCCCCAGTAGGAAAC 872
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db  813 GAGACCACACAATAGTTCACATCCTGCTGTGTGGCTGAAGTGCTGCCCCAGTAGGAAAC 812
```

RESULT 9
US-10-027-632-68743
Sequence 68743, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 10827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 68743
LENGTH: 615
TYPE: DNA
ORGANISM: Human
US-10-027-632-68743

```
Query Match             98.7%;  Score 598.2;  DB 13;  Length 615;
Best Local Similarity   98.7%;  Pred. No. 1.2e-133;
Matches  609;  Conservative   4;  Mismatches   2;  Indels   2;  Gaps   1;

Qy  393 AAGCTTGGTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AAGCTTGGTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy  453 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCCATATTCTTTGCTTTTTTAGTCCTAAAAA 512
        |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||:
Db   61 TTTACCTTCCTGGTATACACCATCCTCCAGCCCCCATATTCTTTGCTTTTTTAGTCCTAAAAM 120

Qy  513 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGAGGGA 572
        |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  121 TCAATAACTGAACTCTCATTGAGTCTAGGCCATTGTGAGTAAACAATAAAGAAGGAGGAGGGA 180

Qy  573 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCAGCCTGGGGCTG 632
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCAGCCTGGGGCTG 240

Qy  633 AGCCTTGGCCTACATCCTGCCCCAAGTGGGAGGATCAGTGCCCCCATTTAAACATCTGGTAGA 692
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AGCCTTGGCCTACATCCTGCCCCAAGTGGGAGGATCAGTGCCCCCATTTAAACATCTGGTAGA 300

Qy  693 CTAAAGAACGCAAGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  693 CTAAAGAACGCAAGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752

Qy  993 CTCTGCCTCTCCAAGCTT 1009
        ||||||||||||||||||
Db  599 CTCTGCCTCTCCAAGCTT 615

Qy  873 AGTTCTTCTAAAGCCTAATAGTCACAGATTCAACAATACCTTCCAGATGTGAAGTGTTAAGCAATTTACAAATTTAAG 932
        |||||||||||||||||||||||||||||||||||||||||||||||         |||||||||||||:||||||||
Db  481 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCAAACTTTTTTGCCAGTTTTCTCTGATATCCCAATCTATTCTTTA 538

Qy  933 GAACTTAAAATAGCCTTCAAAACTTTTTGCCAGTTTCTCTGATATCCAAATCTATTCTTTTA 992
        |||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||:
Db  539 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAAATCTATTCTTTTA 598
```

RESULT 10
US-10-027-632-68744

; Sequence 68744, Application US/10027632
; Publication No. US2002019837lA1
; Publication No. US2002019837lA1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 68744
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68744

Query Match   98.7%;   Score 598.2;   DB 13;   Length 615;
Best Local Similarity 98.7%; Pred. No. 1.2e-133;
Matches 609; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

Qy   393 AAGCTTCGGTGCAAAAGAAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGCTTCGGTGCAAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy   453 TTTACCTTCCTCGGTATACAGCATCCTCCAGCCCCATATCTTCTTTTTAGTCCTAAAAA 512
          ||||||||:|||||||||||||||||||||||||||||||||||||||||||||:
Db   61 TTTACCTTCCTCGGTATACAGCATCCTCCAGCCCCATATCTTGCTTTTAGTCCTAAAAM 120

Qy   513 TCAATAAACTGAACTCTCCATTCATTGAGTGCTAGCCGCCATTGTAGTAGTAAACAAATAAAGAAGGAGGAGGGA 572
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCAATAACTGAACTCTCCATTCATTGATGTCTAGCCGCCATTGTAGTAGTAAACAAATAAAGAAGGAGGAGGGA 180

Qy   573 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGTGCAAAGCACAGCAGCCTGGGGCTG 632
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCTTCTGACAACTGAGAGGAGAGAAATTGTCATCTGAAGTGGTGGAAAGCACAGCAGCCTGGGGCTG 240

Qy   633 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCATTTAAACATCTGGTAGAA 692
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAAYATCTGGTAGAA 300

Qy   693 CTAAAGAAGGCAAGGACCCTGCCACAAATGACTTATTTCCCTGCATTTGATCACGGTCAATCCT 752
          |||||||||::|||::|||||||||||||||||||||||||||||||||||||||||||
Db   301 CTAAAGAAGGCAACKCCCGCCACAAATGACTTATTTCCCTGCATTTGATACGGTCAATCCT 360

Qy   753 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGGAAAAATTGAAATTTACCTA 812
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGGAAAAATTGAAATTTACCTA 420

Qy   813 GAGACCACACATAGTTCACATCCTGCTGTGTGGGCCCCCCAGTAGGAGAAAC 872
          |||||||||||||||||||||||||:||||||||||||||||||||||
Db   421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCCCCCCAGTAGGAGAAAC 480

Qy   873 AGTTCTTCTAAAGCCTATTGTCAACAAATACCTTCCAGATGTTAGCATTTTACAATTAAG 932
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGTTCTTCTAAAGCCTATTGTCAACAAATACCTTCCAGATGTTAGCATTTTACAATTAAG 538

Qy   933 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAATCTATTCTTTTA 992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAATCTATTCTTTTA 598

Qy   993 CTCTCCTCCCAAGCTT 1009
          |||||||||||||||
Db   599 CTCTGCCTCTCAAGCTT 615

RESULT 11
US-10-027-632-70068

; Sequence 70068, Application US/10027632
; Publication No. US2002019837lA1
; Publication No. US2002019837lA1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 70068
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70068

Query Match   98.7%;   Score 598.2;   DB 13;   Length 615;
Best Local Similarity 98.7%; Pred. No. 1.2e-133;
Matches 609; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

Qy   393 AAGCTTCGGTGCAAAAGAAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGCTTCGGTGCAAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

(first alignment block — continuation, RESULT 12)

```
Qy    453 TTTACTTCCTTGGGTATACAGCATCCTCAGCCCCCATATCTTTGCTTTTAGTCCTAAAAA 512
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     61 TTTACCTTCCTTGGGTATACAGCATCCTCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy    513 TCAATAACTGAAACTCTCATTGATGTCTAGGCCCATTGTAGTAAACAATAAAGAAGGAGGA 572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCAATAACTGAAACTCTCATTGATGTCTAGGCCCATTGTAGTAAACAATAAAGAAGGAGGA 180

Qy    573 GGCTTCTGACAACTGAGAGGAAAATTGTCATCTGAAGTGGTCAAGCACAGCCTGGGGGCTG 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCTTCTGACAACTGAGAGGAAAATTGTCATCTGAAGTGGTCAAGCACAGCCTGGGGGCTG 240

Qy    633 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCAGTTTAACATCTGGTAGAA 692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCAGTTTAACATCTGGTAGAA 300

Qy    693 CTAAAGAAACGCACGCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752
          ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db    301 CTAAAGAAACGCACGCTGCCACAAYGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy    753 TGAGAAAATGTTTTCTTTTGGTTTCTCCCTGACAAAGGTTGGGAAAAATTTGAAATTTACCTA 812
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGAGAAAATGTTTTCTTTTGGTTTCTCCCTGACAAAGGTTGGGAAAAATTTGAAATTTACCTA 420

Qy    813 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCAGTAGGAAAC 872
          |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db    421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCAGTAGGAAAC 480

Qy    873 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTT--CATTTAAG 932
          |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db    481 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTT--CATTTAAG 538

Qy    933 GAACTTAAAATAGCCTTCAAAACTTTTTCTCTGATATCCAATCTATTCTTTTTA 992
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 GAACTTAAAATAGCCTTCAAAACTTTTTCTCTGATATCCAATCTATTCTTTTTA 598

Qy    993 CTCTGCCTCTCCAAGCTT 1009
          ||||||||||||||||||
Db    599 CTCTGCCTCTCCAAGCTT 615
```

RESULT 12
US-10-027-632-70069
Sequence 70069, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SEQ ID NO: 70069
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
; SOFTWARE: FastSEQ for Windows Version 4.0

RESULT 13

US-10-027-632-70069

Query Match 98.7%;   Score 598.2; DB 13;   Length 615;
Best Local Similarity 98.7%;
Matches 609; Conservative 4; Pred. No. 1.2e-133; Mismatches 2; Indels 2; Gaps 1;

```
Qy    393 AAGCTTGGTGCAAAGAAATTGGACACATTTCCCAAAAAGTAAGACATACTGGGAAGTCCCTG 452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCTTGGTGCAAAGAAATTGGACACATTTCCCAAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy    453 TTTACTTCCTTGGGTATACAGCATCCTCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAA 512
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db     61 TTTACCTTCCTTGGGTATACAGCATCCTCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy    513 TCAATAACTGAAACTCTCATTGATGTCTAGGCCCATTGTAGTAAACAATAAAGAAGGAGGA 572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCAATAACTGAAACTCTCATTGATGTCTAGGCCCATTGTAGTAAACAATAAAGAAGGAGGA 180

Qy    573 GGCTTCTGACAACTGAGAGGAAAATTGTCATCTGAAGTGGTCAAGCACAGCCTGGGGGCTG 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCTTCTGACAACTGAGAGGAAAATTGTCATCTGAAGTGGTCAAGCACAGCCTGGGGGCTG 240

Qy    633 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCAGTTTAACATCTGGTAGAA 692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCAGTTTAACATCTGGTAGAA 300

Qy    693 CTAAAGAAACGCACGCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752
          ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db    301 CTAAAGAAACGCACGCTGCCACAAYGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy    753 TGAGAAAATGTTTTCTTTTGGTTTCTCCCTGACAAAGGTTGGGAAAAATTTGAAATTTACCTA 812
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGAGAAAATGTTTTCTTTTGGTTTCTCCCTGACAAAGGTTGGGAAAAATTTGAAATTTACCTA 420

Qy    813 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCAGTAGGAAAC 872
          ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db    421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCAGTAGGAAAC 480

Qy    873 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTT--CATTTAAG 932
          |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db    481 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTT--CATTTAAG 538

Qy    933 GAACTTAAAATAGCCTTCAAAACTTTTTCTCTGATATCCAATCTATTCTTTTTA 992
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 GAACTTAAAATAGCCTTCAAAACTTTTTCTCTGATATCCAATCTATTCTTTTTA 598

Qy    993 CTCTGCCTCTCCAAGCTT 1009
          ||||||||||||||||||
Db    599 CTCTGCCTCTCCAAGCTT 615
```

RESULT 13
US-10-027-632-70070
Sequence 70070, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70070
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70070

Query Match   7.4%; Score 598.2; DB 13; Length 615;
Best Local Similarity 98.7%; Pred. No. 1.2e-133;
Matches 609; Conservative   4; Mismatches   2; Indels   2; Gaps   1;

Qy   393 AAGCTTGGTGTGCAAAGAAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGCTTGGTGTGCAAAGAAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy   453 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAA 512
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db   61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy   513 TCAATAACTGAACTCTTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCAATAACTGAACTCTTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy   573 GGCTTCTGACAACTGAGAGGAAAATTGTCATCTGAAGTGGTGCAAGCACCTGGGGCTG 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCTTCTGACAACTGAGAGGAAAATTGTCATCTGAAGTGGTGCAAGCACCTGGGGCTG 240

Qy   633 AGCCTTGGCCTACATCCTCCCCAAGTGGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 692
          ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db   241 AGCCTTGGCCTACATCCTCCCCAAGTGGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy   693 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752
          |||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db   301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy   753 TGAGAAATGTTTTCTTTTGTTCTCCCCTGAGCAAAAGGTTGGGAAAAATTTGAAAATTACCTA 812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGAGAAATGTTTTCTTTTGTTCTCCCCTGAGCAAAAGGTTGGGAAAAATTTGAAAATTACCTA 420

Qy   813 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGGTGAGGAAAC 872
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGGTGAGGAAAC 480

Qy   873 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTTACAAATTTAAG 932
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   481 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGAT--CATTTTAAG 538

Qy   933 GAACTTAAAAATAGCCTTCAAAACTTTTTGCCAGTTTCTGATATCCAATCTATTCTTTTA 992
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539 GAACTTAAAAATAGCCTTCAAAACTTTTTGCCAGTTTCTGATATCCAATCTATTCTTTTA 598

Qy   993 CTCTGCCTCCCAAGCTT 1009
          |||||||||||||||||
Db   599 CTCTGCCTCTCAAGCTT 615

RESULT 14
US-10-027-632-70071
; Sequence 70071, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70071
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70071

Query Match   7.4%; Score 598.2; DB 13; Length 615;
Best Local Similarity 98.7%; Pred. No. 1.2e-133;
Matches 609; Conservative   4; Mismatches   2; Indels   2; Gaps   1;

Qy   393 AAGCTTGGTGTGCAAAGAAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGCTTGGTGTGCAAAGAAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy   453 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAA 512
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db   61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy   513 TCAATAACTGAACTCTTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCAATAACTGAACTCTTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy   573 GGCTTCTGACAACTGAGAGGAAAATTGTCATCTGAAGTGGTGCAAGCACCTGGGGCTG 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCTTCTGACAACTGAGAGGAAAATTGTCATCTGAAGTGGTGCAAGCACCTGGGGCTG 240

Qy   633 AGCCTTGGCCTACATCCTCCCCAAGTGGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 692
          ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db   241 AGCCTTGGCCTACATCCTCCCCAAGTGGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy   693 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 752
          |||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db   301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy   753 TGAGAAATGTTTTCTTTTGTTCTCCCCTGAGCAAAAGGTTGGGAAAAATTTGAAAATTACCTA 812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGAGAAATGTTTTCTTTTGTTCTCCCCTGAGCAAAAGGTTGGGAAAAATTTGAAAATTACCTA 420

Qy   813 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGGTGAGGAAAC 872
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGGTGAGGAAAC 480

Qy   873 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATTTTACAAATTTAAG 932
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   481 AGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGAT--CATTTTAAG 538

Qy   933 GAACTTAAAAATAGCCTTCAAAACTTTTTGCCAGTTTCTGATATCCAATCTATTCTTTTA 992
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539 GAACTTAAAAATAGCCTTCAAAACTTTTTGCCAGTTTCTGATATCCAATCTATTCTTTTA 598

Qy   993 CTCTGCCTCCCAAGCTT 1009
          |||||||||||||||||
Db   599 CTCTGCCTCTCAAGCTT 615

RESULT 15
US-10-027-632-295009
; Sequence 295009, Application US/10027632
; Publication No. US20020198371A1

```
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SEQ ID NO 295009
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-295009
```

```
Query Match             7.4%;  Score 598.2;  DB 13;  Length 615;
Best Local Similarity  98.7%;  Pred. No. 1.2e-133;
Matches  609;  Conservative  4;  Mismatches  2;  Indels  2;  Gaps  1;

Qy    393 AAGCTTGGTGCAAAGAAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 452
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCTTGGTGCAAAGAAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG  60

Qy    453 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAA 512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db     61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy    513 TCAATAACTGAACTTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCAATAACTGAACTTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy    573 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAAGCACAGCCTGGGGCTG 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCTTCTGACAACT-GAGAGGAAATTGTCATCTGAAGTGGTGCAAAGCACAGCCTGGGGCTG 240

Qy    633 AGCCTTGGCCTACATCCTTGCCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTGAGAA 692
          ||||||||||||||:||||||||||||||||||||||||||||||||||:|:|||
Db    241 AGCCTTGGCCTACATCCTTGCCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTGAGAA 300

Qy    693 CTAAAGAAGGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACGTCAATCCT 752
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTAAAGAAGCCCAACKCCTCGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy    753 TGAGAAATGTTTTCTTTGTTCTCCCCTGAGCAAAGGTTGGAAAAATTGAAATTACCTA 812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGAGAAATGTTTTCTTTGTTCTCCCCTGAGCAAAGGTTGGAAAAATTGAAATTACCTA 420

Qy    813 GAGACCACACTAGTTCACATCCTCGTGTGTTGGGCTGAATGTCTGCCCCCCAGTAGGAAAC 872
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAGACCACACATAGTTCACATCCTCGCTGTTGGGCTGAATGTCTGCCCCCYCAGTAGGAAAC 480

Qy    873 AGTTCTTCTAAAGCCTATTGTCAACAAATACCTTCCAGATGTTAGCATTTTACAATTTAAG 932
          ||||:|||:|||||||||||||||||||:|||||||||||||||||||||||||||||||
Db    481 AGTTCTCTAAAGCCTATTGTCAACAATACCTTCAACAATGTTAGCATTTT--CATTTAAG 538

Qy    933 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGAATATCCAATCTATTCTTTTA 992
          |||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    539 GAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAATCTATTCTTTTA 598
```

Qy   993 CTCTCGCCTCCCAAGCTT 1009
          |||||||| ||||||||
Db   599 CTCTCGCCCTCTCAAGCTT 615

Search completed: September 12, 2004, 01:42:10
Job time : 2308 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: September 11, 2004, 10:11:48 ; Search time 12097 seconds
(without alignments)
19849.698 Million cell updates/sec

Title: US-09-939-209A-3_COPY_4110_12150
Perfect score: 8041
Sequence: 1 caattgaagtcctgggcac.........ctaggatcctcaggtggatcc 8041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs; 1493109.0276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database :   EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: em_gss:*
10: gb_est1:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estom:*
16: em_estfun:*
17: em_estfun:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_nus:*
25: em_gss_pro:*
26: em_gss_rod:*
27: em_gss_phg:*
28: em_gss_vrl:*
29: gb_gss1:*
29: gb_gss2:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	344.8	4.3	578	14	CB294770	CB294770 12B22014
2	272.8	3.4	298	14	CD511718	CD511718 AGENCOURT
3	268.4	3.3	864	28	AQ745626	AQ745626 HS_2273_A
4	262.8	3.3	787	12	BI547160	BI547160 60319040S

	No.	Score	%			ID	Description
c	5:	240.4	3.0	647	29	AG090278	AG090278 Pan trogl
c	6:	226.4	2.8	712	28	AQ274674	AQ274674 RPCI-6-11
	7:	194.8	2.4	474	9	AW102832	AW102832 xd38h03.x
	8:	189.2	2.4	734	28	AQ282818	AQ282818 RPCII-89
c	9:	181.2	2.3	852	29	CC904986	CC904986 t0181lba
c	10:	179.2	2.2	437	28	AQ472918	AQ472918 CTBI-E1-
c	11:	172.4	2.1	777	12	BG699761	BG699761
	12:	170	2.1	593	12	BG699541	BG699541
	13:	170	2.1	624	12	BI548918	BI548918
	14:	170	2.1	680	12	BI548997	BI548997
	15:	170	2.1	720	12	BI545493	BI545493
	16:	170	2.1	833	12	BI551559	BI551559
	17:	170	2.1	834	12	BI547188	BI547188
c	18:	169.8	2.1	482	12	AQ457908	AQ457908
	19:	169.4	2.1	803	12	BI551703	BI551703
c	20:	169.4	2.1	454	12	AQ466676	AQ466676 HS_5161_B
c	21:	168.4	2.1	727	14	CB998494	CB998494 AGENCOURT
	22:	168.4	2.1	902	12	BG708958	BG708958 60267510A
	23:	168.2	2.1	760	12	BI550370	BI550370 60319242I
	24:	168	2.1	174	9	AU076999	AU076999
	25:	168	2.1	568	12	AU280510	AU280510
	26:	168	2.1	574	14	CB155278	CB155278 K-BST0213
	27:	167.4	2.1	958	12	BI552207	BI552207 60319256
	28:	166.8	2.1	618	12	BG708946	BG708946 60267480J
	29:	166.8	2.1	778	12	BI666696	BI666696 60329191A
	30:	166.8	2.1	845	12	BI668181	BI668181 60329572J
c	31:	166.2	2.1	521	28	AQ316106	AQ316106 RPCII-10
	32:	166.2	2.1	674	9	AL706644	AL706644 DKFZp586C
	33:	166.2	2.1	912	12	BI546499	BI546499 60319150S
	34:	165.8	2.1	441	28	AQ737144	AQ737144 HS_2247_B
c	35:	165.6	2.1	568	28	AQ757067	AQ757067 HS_5448_B
	36:	165.6	2.1	549	29	CB636174	CB636174 tigr-gss-
	37:	165.2	2.1	712	29	AG181600	AG181600 Pan trogl
	38:	163.2	2.1	490	29	AI024209	AI024209 ov1c10.s
c	39:	162.8	2.0	643	29	AG162554	AG162554 Pan trogl
	40:	162.4	2.0	391	9	AA932951	AA932951 oooo607.s
	41:	162.2	2.0	646	28	AQ535292	AQ535292 RPCI-11-3
c	42:	160	2.0	487	28	AQ211630	AQ211630 HS_3244_A
	43:	159.8	2.0	618	28	AQ285845	AQ285845 RPCII-81
c	44:	159.6	2.0	548	28	AQ393760	AQ393760 CTBI-E1-
c	45:	159.6	2.0	775	28	AQ895749	AQ895749 HS_3176_B

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
CB294770
LOCUS   CB294770   578 bp   mRNA   linear   EST 28-FEB-2003
DEFINITION 12B22014_rev_1_E11_r_087.ab1 Chimpanzee brain library Koos Pan
            troglodytes cDNA clone 12B22014_rev_1_E11_r_087.ab1 5', mRNA
            sequence.
ACCESSION   CB294770
VERSION   CB294770.1   GI:28520200
KEYWORDS   EST.
SOURCE   Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 578)
AUTHORS   Hellmann,I., Zollner,S., Enard,W., Ebersberger,I., Nickel,B. and
            Paabo,S.
TITLE   Selection on human genes as revealed by comparisons to chimpanzee
            cDNA
JOURNAL   Genome Res. (2003) In press
COMMENT   Contact: Paabo S
            Evolutionary Genetics
            Max-Planck-Institute for evolutionary Anthropology
            Deutscher Platz 6, 04103 Leipzig, Germany
            Tel: +49-(0)-341-3550 500
            Fax: +49-(0)-341-3550 555
            Email: paabo@eva.mpg.de
            Seq primer: M13 reverse.

FEATURES   Location/Qualifiers
   source   1..578
     /organism="Pan troglodytes"
     /mol_type="mRNA"
     /db_xref="taxon:9598"
     /clone="12B2014_rev_1_E11_r_087.ab1"
     /sex="male"
     /tissue_type="brain, presumably cortex"
     /dev_stage="adult"
     /lab_host="Epicurian Coli (TM) XL-10-Gold"
     /clone_lib="Chimpanzee brain library Koos"
     /note="Vector: pUChi; Site 1: SfiI-A; Site 2: SfiI-B; The
     library was prepared using the SMART cDNA library
     construction Kit (Clontech), doing only primer extension,
     but not PCR amplification of the cDNA. The only deviation
     from the published protocoll was that we cloned the cDNA
     into a plasmid Vector."

ORIGIN

Query Match   4.3%;   Score 344.8;   DB 14;   Length 578;
Best Local Similarity 99.4%;   Pred. No. 5.3e-58;
Matches 346;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;

```
Qy   5606 TTACAGGACCGCAGGCCATGTGAAGGAGGAAACTGTCAGTGCTAGGCTGAATTCTGTTGTTAC 5665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTACAGGACCGCAGGCCATGTGAAGGAGGAAACTGTCAGTGCTAGGCTGAATTCTGTTGTTAC 60

Qy   5666 CAAGATTTCTAGAAAAAGTATTCCTCAGTCAGGTTGATTATACAGATATAGCCAAAATCTATTT 5725
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAAGATTTCTAGAAAAAGTATTCCTCAGTCAGGTTGATTATACAGATATAGCCAAAATCTATTT 120

Qy   5726 TCCTAGGGTAGTTCTTTCGTTATGCTCGCCGGGGCCTTATATAAACTGTCTCGTCATCCAGTTATTTCTC 5785
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCCTAGGGTAGTTCTTTCGTTATGCTCGCCGGGGCCTTATATAAACTGTCTCGTCATCCAGTTATTTCTC 180

Qy   5786 TCCACCTTCTTCTTGTTTTCATAACACCAAGCGCAACTTCCGCCAAATCACTCTGCCGTGGGACGA 5845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCACCTTCTTCTTGTTTTCATAACACCAAGCGCAACTTCCGCCAAATCACTCTGCCGTGGGACGA 240

Qy   5846 TGATCCTGCCAGCTCCCCTTTGGAAATCCTGAGGGATCAGATCTTGGACCATGTATATAATAT 5905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGATCCTGCCAGCTCCCCTTTGGAAATCCTGAGGGATCAGATCTTGGACCATGTATATAATAT 300

Qy   5906 GATGCTTCTAATCCAAAAGAGGAAAAGGCATTGGGAGTGGAGTGCAGCTCCTAAG 5953
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GATGCTTCTAATCCAAAAGAGGAAAAGGCATTGGGAGTGGAGTGCAGCTCCTAAG 348
```

RESULT 2
CDS11718
CDS11718/c
LOCUS   CDS11718   298 bp   mRNA   linear   EST 06-JUN-2003
DEFINITION   AGENCOURT 14360795 NIH_MGC_187 Homo sapiens cDNA clone
   IMAGE:30405362 5', mRNA sequence.
ACCESSION   CDS11718
VERSION   CDS11718.1   GI:31443436
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
   ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 298)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Daniela S. Gerhard, Ph.D.
   Office of Cancer Genomics
   National Cancer Institute / NIH
   Bldg. 31 Rm10A07 Bethesda, MD 20892
   Email: cgapbs-r@mail.nih.gov
   Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
   cDNA Library Preparation: CLONTECH Laboratories, Inc.
   cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
   DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM192   row: g   column: 03
High quality sequence stop: 298.

FEATURES   Location/Qualifiers
   source   1..298
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:30405362"
     /lab_host="DH10B (T1 phage-resistant)"
     /clone_lib="NIH_MGC_187"
     /note="Organ: Blood vessels - aorta, basilar and artery;
     Vector: pDNR-LIB; Site 1: SfiI (ggccatatggcc); Site 2:
     SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in
     cloning as follows: 5' adaptor sequence:
     5'-CACGGCCATTATGGCC-3', and 3' adaptor sequence:
     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
     (range 0.5-4.0 kb). 14/15 colonies contained inserts by
     PCR. This library was enriched for full-length clones and
     was constructed by Clontech Laboratories (Palo Alto, CA).
     Note: this is a NIH_MGC library."

ORIGIN

Query Match   3.4%;   Score 272.8;   DB 14;   Length 298;
Best Local Similarity 97.5%;   Pred. No. 1.2e-43;
Matches 277;   Conservative 0;   Mismatches 7;   Indels 0;   Gaps 0;

```
Qy   2791 TTTTTTTTTTTAAATTTTTAAAAATTTGTATATATTTATGGGGTTATAAGTGAAAACTTTTTAGA 2850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    286 TTTTTTTTTTTAAATTTTTAAAAATTTGGTATATATTTATGGGGTTATAAGTGAAAACTTTTTAGA 227

Qy   2851 TGCATAGGTTGTATAGTGATAAAAATCAGGGCTTTAGGGTATTCATCTACCTGAATGATGT 2910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    226 TGCATAGGTTGTATAGTGATAAAAATCAGGGCTTTAGGGTATTCATCTACCTGAATGATGT 167

Qy   2911 ACATTGTACCCCTTAAGTAAATTTCTCACCATCCGCGCGACTTCTTGCCCCCTGGGTATTCA 2970
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    166 ACATTGTACCCCTTAAGTAAATTTCTCACCATCCGCGCGACTTCTTGCCCCCTGGGTATTCA 107

Qy   2971 TCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTCTCACCATCCGCTGACTTCTTG 3030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    106 TCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTCTCACCATCCGCTGACTTCTTG 47

Qy   3031 CCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTT 3074
          |||||||||||||||||||||||||||||||||||||||||||
Db     46 CCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTT 3
```

RESULT 3
AQ745626
AQ745626/c
LOCUS   AQ745626   864 bp   DNA   linear   GSS 16-JUL-1999
DEFINITION   HS_2273_A2_B06_T7C CIT Approved Human Genomic Sperm Library D Homo
   sapiens genomic clone Plate=2273 Col=12 Row=C, genomic survey
   sequence.
ACCESSION   AQ745626
VERSION   AQ745626.1   GI:5523148
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
   ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 864)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
   Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
   Hood,L.
TITLE   Sequence-tagged connectors: A sequence approach to mapping and
   scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
PUBMED   10449764

FEATURES   Location/Qualifiers
     source   1..864
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2273 Col=12 Row=C"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
                     High quality sequence stop: 864.

ORIGIN

Query Match   3.3%; Score 268.4; DB 28; Length 864;
Best Local Similarity 70.7%; Pred. No. 8e-43;
Matches 454; Conservative   0; Mismatches 156; Indels 32; Gaps   6;

Qy   1139 TTCATTTATTCTTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAAACGTGTGATACATTT 1198
                ||| ||||| | ||| ||| || |||| || ||| |||||| |||||||||| ||||
Db   703 TTTAAATATTGTTTCTGAATTCCATTTAGTTGTTAACGGGCTGAAAGGTGTGATATCTTTT 644

Qy   1199 TCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAACAGCAGGTTAATAAGAG 1258
                |||| ||| | || ||| | |||| |||| ||| |||| || || || | |||
Db   643 CCTCACCCGTTATAATCATCACAGCCAACACTCCTATAACAAAACCAGGATAACAAGAA 584

Qy   1259 CAAAAACCTAACAAATTTATTTATTTACATGACATGGGAGTGTCTTCAGAAATGA 1318
                | ||| | ||||| | |||||||||| || | ||||| ||||||||| | ||
Db   583 AAA-------AACATATTTCTTTAATCAAACTTTACATAGCATAGGAGGCTTCAGAAATGA 529

Qy   1319 AGACCCAAAGACACCAGGGGAAACTGTCTCGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGA 1378
                | |||||| ||| | || |||| ||||| ||||||||| || || || || |||||| |||
Db   528 AAACCTAAAAATCAGGGAAAAATAATTTG-TTTTTATACTTAGGTTCAATGAAGAATAGA 470

Qy   1379 TAGCATGTAGCCATGTAGATTAGAGACAAAAAGGATATGATCCTACTAGTGGTAAAGGACTCAGGG- 1437
                || |||| || || ||| ||| | | ||| |||| ||||| | | || || | ||||| ||||
Db   469 CAGCTGTGTAGAAATGTGATTAGCACAAAAGGGTATGATCTAAAGTAATGGTAACAGACTGACGGA 410

Qy   1438 GGAAACACAGGACAAGGCCTGTGTCTATTCAGATTCAGATTCTCTCTCCTATGTGTAATAG 1497
                || | | ||| ||| ||| | | |||||| || |||| ||| ||||| ||| | |||
Db   409 GGAGATCCAGCAAAGCCTGCTTGT-CAGATTCTTCTTGGCCTCTCGT-------GCA 359

Qy   1498 CATTCTTTCCCTGAGTATGGGGCCAGGACTCTCTTCTTCAATGAGGTCTTCAAGGGACA 1557
                |||| || | | ||||| |||| ||| ||| ||| | | |||||| ||| | || |||
Db   358 ACATTCTTCCCCCAGATTTGGGGCCAGGATCCCTTATGGAATGAGGAGAGAGAA 299

Qy   1558 GGGAGAA-------AGTGGGCTTTTTAGATTTTATGGCTTGCTTCGGGGGAGGAGAGTTC 1610
                ||||||| ||| ||||| ||| ||| |||| ||| ||| || |||||||| ||
Db   298 TGGAAAAGGGAAAGAGTAGCCTTTCTAGGATTGATGGCTTCGTTGGGGGAGGAGAGATTC 239

Qy   1611 TAGTTTCTAATGACCCATCTTGGGGAAGGAGGAATTCTGGGTTTCTGTGACTTGCTTTCAATGA 1670
                ||| ||||| | |||| ||| || || ||| |||||||| | || |||| || | |||
Db   238 TAGTTTCTATGACCCACATTAGGGGAAATGAATTCTGGGTCTCTATGACTTCCGCTTCAGGGG 179

Qy   1671 AGAAAGAGGAGTAAGAGGCAGGAGGCCAGGAGGAGTGGTCAGAAGAGACT--------TG 1721
                ||||| || || | ||| || ||| || | || || || ||| | ||| ||
Db   178 AGAAGGAGGGGGGCAGCAGGAGGAGGCAGAAGAGAGGCTCATGCTTCTGG 119

Qy   1722 GCTCGTTCTGAGGGCCTTCCGCTCTCCTTTAGTTCCAAGTACT 1763
                ||| || |||| | ||||| | || || ||||||| |||
Db   118 GCTCGTTTCTGAGGCCTTCAGCTCTCCTTTAGCTCAAAGTACT 77

───

RESULT 4
BI547160
LOCUS   BI547160   787 bp   mRNA   linear   EST 05-SEP-2001
DEFINITION 6031904O5F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5261664 5',
            mRNA sequence.
ACCESSION   BI547160
VERSION   BI547160.1 GI:15434472
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 787)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            http://image.llnl.gov
            Plate: LLAM1659 row: j column: 01
            High quality sequence stop: 721.

FEATURES   Location/Qualifiers
     source   1..787
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5261664"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI), National Institutes of Health). Note: this
                     is a NIH_MGC Library."

ORIGIN

Query Match   3.3%; Score 262.8; DB 12; Length 787;
Best Local Similarity 99.2%; Pred. No. 1.1e-41;
Matches 264; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   6223 GCAGAGGGGAGACAGAGGAGCTGGGACTGGATCTGAGAGAGGTCGTCTGATTGGTGGACGGTCGT 6282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GGAGGGGGGAGACAGAGGAGCTGGTACTGAGAGAGGTCGTCTGATTGGTGGACGGTCGT 63

Qy   6283 AGCTGGGCTATAAAAGAGACCCCTACAGGGTTAGCGAGGAGAGACCCTCAGAGGATTCTGAC 6342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 AGCTGGGCTATAAAAGAGACCCCTACAGGGTTAGCGAGGAGAGACCCTCAGAGGATTCTGAC 123

Qy   6343 AATATCTTTACCGGAGAGAGGGGCAAAGTACGGCTCAAAAGCCGAAGCCACAGCTCCTCCCGC 6402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 AATATCTTTACCGGAGAGAGGGCAAAGTACGGCTCAAAAGCCGAAGCCACAGCTCCTCCTGC 183

Qy   6403 CGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAGATGTGCAAAGGGCCTTGCA 6462
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAGATGTGCAAAGGGCCTTGCA 243

Qy   6463 GGTCTGCCGGCGCTTCCTTGCTTGAGGAG 6488
              |||||||||||||||||||||||||||||
Db   244 GGTCTGCCGGCGCTTCCTTGCTTGAGGAG 269

RESULT 5

```
AG090278
LOCUS        AG090278/c
DEFINITION   Pan troglodytes DNA, clone: PTB-089M15.R, genomic survey sequence.
ACCESSION    AG090278
VERSION      AG090278.1  GI:16642080
KEYWORDS     GSS.
SOURCE       Pan troglodytes (chimpanzee)
ORGANISM     Pan troglodytes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE    1  (bases 1 to 647)
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        BAC end sequences of Library PTB
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 647)
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
             and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
             1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
             Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      Clones are derived from the chimpanzee BAC library PTB This BAC end
             was generated during the R&D process and may have higher chance of
             clone tracking errors.
             PRIMERS
             Sequencing: M13Rev
FEATURES             Location/Qualifiers
LIBRARY
             Vector     : pKS145
             R.Site 1   : SacI
             R.Site 2   : SacI.
             source     1..647
                        /organism="Pan troglodytes"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9598"
                        /clone="PTB-089M15.R"
                        /sex="male"
                        /cell_type="lymphoblast"
                        /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
```

Query Match   3.0%; Score 240.4; DB 29; Length 647;
Best Local Similarity   69.3%; Pred. No. 3.1e-37;
Matches 424; Conservative 0; Mismatches 161; Indels 27; Gaps   6;

```
Qy    1194 CATTTTCTTCCATCATAAGGGTCGGCACCAAAAACTCCTATAGTAAAAGAGCAGGGTTAAT 1253
                |||| | | ||||||||| || ||||     ||   |||  || ||| |||||
Db     647 CCTTTCCTCACCCATCATAAGGGTCACGACTGACATTTTATAACAAAAGACAGGTTAAC 588
Qy    1254 AAGAGCAAAAACCTAACAAATTTATTTAATCAAAAGTTTTACATGCATGGGGAGTCTTCAGA 1313
               ||||||     ||  ||   ||||||||||||||   |||| ||     ||||||||||
Db     587 CAGAGAAAAGCATAACAAATTTATTTAATTGAAGTTGTCATCACACATGAGCCTTTTAGG 528
Qy    1314 AATGAAGACCCAAAGACCCAGGGGAAACTGCTGTTTTTTTTTTTCCTGAGGTTCGGATGAAGA 1373
            || ||||| |||||| ||||| ||||  |||||||||||||||||||||   |||||||||||
Db     527 ATGAGAGACCCAAGA---CCTGGGAAAACT---ATTTTTATGCTTAGGTTCAGTGAAGA 476
Qy    1374 ATGGATAGCATGTAGCCATGTAGATTAGAATAGACAAAAGGATATAGATCTAGTGGTTAAAGGACTC 1433
             |||||||| || |||||| || || ||  ||| ||||||||||||||||| |||| |||||||
Db     475 ACAAACGGCCATGTAGAAATGTGGCCAAAAAACTGTGTTCTAATGGTTAAGAGACCGAGTG 416
Qy    1434 AGGGGGAAAACACAGCAAAGGCCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTATGT 1493
             ||| ||| ||||||   |||| ||||| ||||||||||| |||||||||||| ||||||
Db     415 GAGAAGGCCAGCCGCCGGAGGACTGTGTGTTCAGATTCTTCCTGGTTTTTTC-----TGT 364
Qy    1494 ATAGCATTCTTTCTCCCCTAGTATGGGGGTCAGGGACTCTTCCTCAATGAGGGTCTTCAAAGG 1553
            |||||||||||||| |||||||||||||| ||||||| ||| |||||| ||||| |||||||
```

RESULT 6

```
AQ274674
LOCUS        AQ274674             712 bp    DNA     linear   GSS 26-JAN-2001
DEFINITION   RPCI-6-112A10Sp6 RPCI-6 Homo sapiens genomic clone
             RPCI-6-112A10Sp6, genomic survey sequence.
ACCESSION    AQ274674
VERSION      AQ274674.1  GI:3851781
KEYWORDS     GSS.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 712)
AUTHORS      Beck,A., Cairo,A., Catanese,J., Osoegawa,K., van Geel,M., Wu,C. and
             de Jong,P.J.
TITLE        BAC and PAC end sequence database for sequence-ready map building
JOURNAL      Unpublished (1998)
COMMENT      Contact: de Jong, P.J.
             Children's Hospital Oakland Research Institute
             747 Fifty second Street, Oakland, CA 94609-1809, USA
             Tel: 510 450 7911
             Fax: 510 450 7924
             Email: pdejong@mail.cho.org
             PAC end sequences. For clone availability please contact Pieter de
             Jong (pdejong@mail.cho.org). BACPAC Resources WWW site:
             www.chori.org/bacpac
             Seq primer: Sp6
             Class: PAC end.
FEATURES             Location/Qualifiers
             source     1..712
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /clone="RPCI-6-112A10Sp6"
                        /sex="female"
                        /cell_type="lymphocytes"
                        /lab_host="E. coli DH10B"
                        /clone="RPCI-6"
                        /clone_lib="RPCI-6"
                        /note="Vector: pPAC4; PAC clones in E. coli DH10B. For PAC
                        library availability, please contact Pieter de Jong
                        (BACPACorders@mail.cho.org). Clones may be purchased from
                        BACPAC Resources (www.chori.org/bacpac) or from Research
                        Genetics (info@resgen.com)"
ORIGIN
```

Query Match   2.8%; Score 226.4; DB 28; Length 712;
Best Local Similarity   71.5%; Pred. No. 1.9e-34;
Matches 418; Conservative 0; Mismatches 137; Indels 30; Gaps   8;

```
Qy    1176 GGCTGAAAACGTTCTGATACATTTTTTTGCTTGCCATAAGTTTCCGCTAACCAAAAACCTTAT 1235
Qy    363  GTAGCATTCTTTCCTCCTCGGGTCTGGGGGTGTCGGGCCAGGAGCGCCGCCCCTCGGGAAGGAGGAGTTCTTCAAAGG 304
Db    1554 AGAAGGGGAGAAAGTGGCCTTTTTTGAATTTATTGGTCTTGCTTGCTTCCGGGAGGAGGAGTTCAG 1613
Db    303  AGAAGGGGAGAAGTGGACTCTTTGTAGGTTGTCTATGGCTTGCTTGCTTGGGGAGGA-GAATTCTAG 245
Qy    1674 AAGAGGGAGTAAGAGGCCAGGAGGGGCCAGGAGAGAGTGGTCAGAAAGACTTGGCTCTGCTTCTGAG 1733
Db    184  AAGAGGGGCAAAGAACAGAGAGGGCCAGGGAGGAAAGATCGGAAA------GTTTGCTTCTGAG 132
Qy    1734 GGCTTCCGCTCCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCAACTATACTTTGGC 1793
Db    131  ATCTTCCAGTCTCCTTTAGTTTGAAGTA---TTCGGTGTATCAAAAGCACCAGACTTTCGA 75
Qy    1794 ATATGGTTTTCT 1805
Db    74   GTATTGTCTTCT 63
```

Left column

```
FEATURES             Location/Qualifiers
     source
```

High quality sequence stop: 409.
Seq primer: -40UP from Gibco
Possible reversed clone: polyT not found

www.bio.llnl.gov/bbrp/image/image.html
found through the I.M.A.G.E. Consortium/LLNL at:
Clone distribution: NCI-CGAP clone distribution information can be
DNA Sequencing by: Washington University Genome Sequencing Center
cDNA Library Arrayed by: Greg Lennon, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
Emmert-Buck, M.D., Ph.D.
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
JOURNAL   Unpublished (1997)
Tumor Gene Index
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
REFERENCE   1 (bases 1 to 474)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM   Homo sapiens
SOURCE   Homo sapiens (human)
KEYWORDS   EST.
VERSION   AW102832.1 GI:6073445
ACCESSION   AW102832
sequence.
similar to contains MER40.b2 MER40 repetitive element ;, mRNA
DEFINITION   xd38h03.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2596085 3'
LOCUS   AW102832   474 bp   mRNA   linear   EST   19-OCT-1999
AW102832
RESULT 7
```

### Alignment block (RESULT 7)

```
Db   570 AAGCCTTGGCTTCT--GAGGNCCTTCCATTCTTCCTTTTAGTTCAAA 612
Qy  1714 GAGACTTGGCTGCTTCTGAGGGGCTTCCGCCTCCTCCTTTTAGTTCAA 1758
Db   513 ATGACTTACTTACGTCGGGGAGAAAAGGATTGGGGAGAGAGAAGG---GTGA 569
Qy  1654 GTGACTTGCTTCATGAAGAAAGAGGAGTAAGAGGCCAGGAGGGGCAGGAGGGGTCAGAAA 1713
Db   453 TTGCAGGAGAGGCAATTCTAGTTTCCTACTTTGGCGAAAAAAGAGAATTCTGTTTTCT 512
Qy  1594 TCGGGGAAGAGGAGTTCTAGTTTCTTATGAGGAGGAGGACCGAGTGGAGAAGAAACCACAGCA 1653
Db   393 GGAATGAGGGTTTCAAGCGAGAGGAGGGGACCTTTCTAGGTTTGTGGCTTGCT 452
Qy  1534 TCAATGAGGGTCTTCAAGGAGGAGAAAGTGGCGCCTTTTTTAGATTTATGGCTTGCT 1593
Db   348 TAGTGTCTC-----------TGAGTAGGCATTCCTACCTCC------CAGGTAGGACCTCTCT 392
Qy  1474 TGACTCTCTCTCTCCTATGTATAGCATTCTTCCTCTGAGGTATGGGGCAGGACTCTCTTCT 1533
Db   288 TCTACTGGTAACAGTGAGGGGGAGAAACCCAGGCCCTGTCTCTGTTCAGATTCTTCT 347
Qy  1416 TCTAGTGGTAAAGGACTCAGG--GGAAACAGCAAGGCCTGTCTCATTCAGATTCTTCT 1473
Db   228 GCTTAGTTCCATGAAGAAATGGGGAAGCCTTGTTGAGAAATGTGATTGGACAAAAAGGGGTGA 287
Qy  1356 GCTGGAGGTTGGATGAAGAATTGGCATGTAGCCCATGTAGATTAGACAAAAGGAATATGA 1415
Db   173 GACACAGAAGCCCTTCAGAAATGAGAGACTCAAAGACCAGGGAAAAACTG-----TATTTGT 227
Qy  1296 GACATGGGAGCTCAGAAAATGAGAGAGACCCAAGACGGGGAAAACTGTCTGTTTTTTTT 1355
Db   116 AACAAAAG--AGGTTAACAAGAGAAAAAGCATAAATGCATTTATTTCATC-AGTTCACAT 172
Qy  1236 AGTAAAAGACAGGTTAATAAGACGCAAAACCTAACAAATTTATTTAATCAAAAGTTTTACAT 1295
Db    56 GGGATGGAAAGTGTGATAACCTTTCCTATCTACCATACCATAAGGATTACGGCCAACACTGCTAT 115
```

Right column

```
ORIGIN
     Query Match   2.4%;  Score 194.8;  DB 9;  Length 474;
     Best Local Similarity  73.5%;  Pred. No. 3.8e-28;
     Matches 322;  Conservative 0;  Mismatches 97;  Indels 19;  Gaps 5;

Db    19 AACAAAATTTATTTAAAGTTTTACAITTACATTAGGAG-CCTCAGAAAATGAAGACC--AAAGA 75
Qy  1271 AATTTATTTAATCAAAAGTTTTACAITGACATGGGAGGTCTTCAGAAAATGAAGAACCAAGAC 1330
Db    76 CCCAGGAAAAAATGATCTATTTGTAITGCTTTATGTTTGATGAAGAATGGGACAGCCATATAGA 135
Qy  1331 CCCAGGGAAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAAATGGATAGAGTGTAGCC 1390
Db   136 AATGTGTTGGACAAAAAGGGTACCACTTGGTGAGTAGACTCGCAGGGGAAAACCCAGCA 195
Qy  1391 ATGTAGATTAGACAAAAGGATATGAATCCTAGTGGTAAAAGGACTCAGGGGAAACCACAGCA 1449
Db   196 AGGCCTGTTCGTTCAGAGTTCTTCTGGCCTCTC-------TGTGTAGCAATTCTTTCCTC 247
Qy  1450 AGGCCTGTCTCATTCAGATTCTTCTCTCTCTCTATGTATAGCATTCTTTCCTC 1509
Db   248 CTGGGTGTGGGACAGGACCCCTCTGGAATGAGGGGTCTTCAAGGAGGAAAGGGGCAAAAGGAG 307
Qy  1510 CTGACTATGGGGCAGGACTCTCTTCTTCTTCAATGAGGGTCTTCAAGGGGAG------AAGGGAG 1562
Db   368 CCCGGCCTTGGGGACAGAGAAATTTTGGTTTCTATGCCTCACTTCAGGGGAGAAAGAGAGAGGT 427
Qy  1623 CCCATCTTGGGGAAGAGGAGAATTCTGGGTTCGTGGACTTGCTTTCATGGAAGAAAGAGGAGT 1682
Db   428 GAGAGACAGGAGGGGCGGG 445
Qy  1683 AAGAGGCAGGAGGGGCAGG 1700

RESULT 8
AQ282818
LOCUS    AQ282818    734 bp    DNA   linear   GSS   27-APR-1999
DEFINITION   RCI11-89G8_TV RPCI-11 Homo sapiens genomic clone RPCI-11-89G8,
genomic survey sequence.
ACCESSION   AQ282818
VERSION    AQ282818.1  GI:3909290
KEYWORDS   GSS.
SOURCE    Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 734)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE    Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL   Unpublished (1998)
COMMENT   Other_GSSs: RPCI11-89G8.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
```

The `source` entry (right side):

```
source          1..474
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:2596085"
                /tissue_type="tumor, 5 pooled (see description)"
                /lab_host="DH10B"
                /clone_lib="NCI_CGAP_Ov23"
                /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.35 kb. Tumor types include: mixed
                Mullerian tumor, papillary serous, clear cell, spindle
                cell. All are primary tumors, metastasis positive. Life
                Technologies catalog #: 11534-013".
```

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7

FEATURES   Location/Qualifiers
     source   1..734
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7533943"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-89G8"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"

ORIGIN
     Query Match
     Best Local Similarity   66.0%;   Score 189.2; DB 28; Length 734;
     Matches 289; Conservative   0; Mismatches 148; Indels   1; Gaps   1;

Qy   1050 CTACCTCTTCTCCCTCCATTAACTACAGAGTAAAATTCTCGGTCTTCAGAGTAACAAGAAACA 1109
                ||||||||| ||||||| ||||| | | ||| | || || ||||||||||||| ||||
Db   280 CTAGACTTTCACTTATTTTAATTGATGCACAATGCCTTATTTAAAAGTTTCAGCCTTAA 339

Qy   1110 CCCTTTAGTTCTTCAGCATATTCGTGCACCTTCATTTATCTCCCTCTTCTCTCAAAGCTG 1169
                ||| || ||||| |
Db   340 TCCTCCCCTCCTCCTGCTTTTTTGTGAGATTTTCTGGTCTCATCCAACTACAGTTCTTTCCAG 399

Qy   1170 CAGTAGGGGTGGAAAACGTGTGATACATTTTCTCTTCTTCCATCATAAGGGTCGCAACCAAAAC 1229
               || ||
Db   400 GTTTCGGGGATGGAAAAGCGTGGTACCTTTCATCCCATCATAAGGGTTCACAGCCAACAC 459

Qy   1230 TCCTATAGTAAAAGACAGGTTAATAAGAGCAAAATTTATTTATTTACAAAGTT 1289
               |||||||
Db   460 TCCTATAACAAAAAACAGGTTAAACATGAGAAAAAGCAATAACATTTCTTTAATCAAAAGTT 519

Qy   1290 TTACATGACATGGGAGAGTCTTCAGAAATGAAAATGACCCAAGAGACCCAGGGGAAACTGTCTGTT 1349
               |||||||||| |||| ||| ||||
Db   520 TTATATGACATGGGAGAGCTTCAGAAATGAAAATGAGAACCCAGGGAAAAACTGTCT-AT 578

Qy   1350 TTTTTTGCTCAGGGTTCGGATGAAGAATGATAGCATGAGGGCCCATGTAGTAGCCACAAAAAGG 1409
               ||||| ||||||| |
Db   579 TTTTTTTTGCTTAGGGATCAAAGAAGAATGACAGACGCCATGTAGAAATGTGATTGGACAAAGG 638

Qy   1410 ATATGATCTAGTGTGGTTAAAGGACTCAGGGGGGAAACACAGCCAAGCCCTGTTCTTATTCAGATTC 1469
               ||||||
Db   639 TTGTGATCTAATGGTTGATAGTTGCATGGGAAAACCCAGCAAGCCCTGACTGTTCAGATTC 698

Qy   1470 TTCTTGATCTCTCTCTCT 1487
               ||||||||||||| ||||
Db   699 TTCTTGGCCTCTTTGTGT 716

ORGANISM   Bos taurus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.;
           Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
           Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 852)
  AUTHORS Lin,S. Najar,F.Z., Adelson,D., Gill,C.A. and Roe,B.A.
  TITLE   Bovine BAC End Sequences from Library TAMBT
  JOURNAL Unpublished (2003)
COMMENT   Contact: Bruce A. Roe
           Advanced Center for Genome Technology
           University of Oklahoma Department of Chemistry and Biochemistry
           620 Parrington Oval, Room 208, Norman, OK 73019, USA
           Tel: 405 325 4912
           Fax: 405 325 7762
           Email: broe@ou.edu
           Class: BAC ends
           High quality sequence start: 109
           High quality sequence stop: 403.

FEATURES   Location/Qualifiers
     source   1..852
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="Angus bull T A M U Shoshone Y6 11519666"
                     /db_xref="taxon:9913"
                     /clone="t01812lba"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="TAMBT"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; TAMBT Bovine BAC library (Male) produced by Texas
                     A&M University, Department of Animal Science."

ORIGIN
     Query Match
     Best Local Similarity   65.3%;   Score 181.2; DB 29; Length 852;
     Matches 372; Conservative   0; Mismatches 166; Indels 30; Gaps   6;

Qy   3338 TTGCTCTTCATTTCATTTGGAAATACTAATTCCTAGGCTTCTTGGCATTGGTTGCTTCTCCTTCT 3397
               ||||||| ||||||| ||
Db   622 TGGCTCTTACCCTCTATTTGGTATA-TATTTCGTAAGCTTGTTGTCATTTGGTGGTTGGT 564

Qy   3398 CACCCATC-AAATCTCATTTAGATACACACCTCTTCAAAGAGAGGGCCTTCCTGACCACCTT 3456
               ||
Db   563 CAGCCATCAAAAATCTCATTTCATTTCACTATATCACTTGGTCAAGGAGGGGACTTCC--------- 514

Qy   3457 GGCTGAATTAGCCCTTCACCACTCGATTACTCTCTAGCACATCACCTGCCCATTTATTC 3516
               ||||
Db   513 --CAGATACACCCCTTTGCTATTCAGTTTCTTTGTATCA-------GTTGTCAGTTTTTTT 462

Qy   3517 ATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTTCTGACTCCTTCTACTGA 3576
               ||||||
Db   461 TCATGATTATGGGTGACTCTAGAATTATTTTTATAGTTT-CTCAACTCTTTTTTCTACTGG 403

Qy   3577 GATGAAAAACTCTACTAGAGCGAGAGATTTTATTCTGTCTTGTATCAGGTACTG---------C 3627
               |||||
Db   402 GGTGAAAACTCCACTAGGGTGGGGAACCCGTTTGCTTGCTCATCATACTAACAGACAGTTC 343

Qy   3628 TTCAAACACGACACCTGATACAGAGTAGGTGGTCAAAAGATAATTTCTTAAACAAAATGAACAA 3687
               |||||
Db   342 TTCAAACGGTACCTGTCCCAGAGTAGGAGCTCAAAAATATAAGTATTATTAAATAAGTGAATAA 283

Qy   3688 ATAAAAGTAGATCTTTGAGAGTAGAAAGCTCTTCCACACTACCACGAGTCATTCAGGAATG 3747
               ||||||||
Db   282 ATAAAAATTAAATCTTTTCAGAATAAAAGCTCTTCACATCACCCAGGAATCACCCAGAATG 223

Qy   3748 ACAAATCATAGAATAACAGAAATTGATGCTTGTGTGCATATATCAGAGAAGAAGGTGGAAGG 3807
               ||||
Db   222 ACAGGTCTTAGAATAATAAAGAGTGTAATGCTTTCGCATGTCAGAGAGAAAGAGGGTGGAAAT 163

Qy   3808 TTGTCAAGGTATCATGATGTACCAGTGTCTCGCGCCTCCTCAAACACAATCTGCAAGTCCAC 3867
               ||||||
Db   162 TTGTCAAGGTATCATGATGATGTACCAAATATTTGCCTCCCATAGTAGAAAAATACATTATGAA 103

RESULT 9
LOCUS   CC904986   852 bp   DNA   linear   GSS 08-AUG-2003
DEFINITION t01812lba.f1 TAMBT Bos taurus genomic clone t01812lba, genomic
            survey sequence.
ACCESSION   CC904986
VERSION   CC904986.1 GI:33523919
KEYWORDS   GSS.
SOURCE   Bos taurus (cow)

RESULT 10

```
Qy    3868 AGTGAAAAGTAAGTTAACTTCAATGTGAAGC 3897
Db     102 AGGGTTGCAATGCCTGCAGGTTGATGTGACGC 73
```

```
LOCUS       AQ472918           437 bp    DNA    linear    GSS 23-APR-1999
DEFINITION  CITBI-E1-2589A4.TR CITBI-E1 Homo sapiens genomic clone 2589A4,
            genomic survey sequence.
ACCESSION   AQ472918
VERSION     AQ472918.1  GI:4656484
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Homo sapiens
            Other_GSSs: CITB1-2589A4.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbee@tigr.org
            Clones are available from Research Genetics (info@reggen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..437
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2589A4"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITB1-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     Caltech Human BAC Library D"
ORIGIN
```

```
Query Match     2.2%;  Score 179.2;  DB 28;  Length 437;
Best Local Similarity    72.9%;   Pred. No. 4.9e-25;
Matches   334;  Conservative   0;  Mismatches   98;  Indels   26;  Gaps    7;
```

```
Qy  1186 GTGGTGATACATTTTCTTCCATCAATAAGGTCGCGAACCAAAAACTCCTATAGTAAAAGAC 1245
         ||||  |||    |||||| ||| ||  |||||    ||||| ||||||||  || |||||
Db   436 GTGTGATACCTTCATCAGCCATCATTAAGG------GCTAACACTCCTATAACAAAAAACA 382

Qy  1246 AGGTTAATACAGACAAAAACCTAACAAAATTTATTTTAATTTTACATGACATGGGAG 1305
         ||||| |||  |||||||||      ||||||||   ||| |||||||||||||||||
Db   381 AGTT--ATCGAGAAAAACATAACACATTTATTTAATCAGAGTTTTACATGGCACAGGAG 325

Qy  1306 TCTTCAGAAAATGAACAGACACCCAAAGACCCAGGGAGGAAACTGTTCGTTGTTTTTT 1365
         |||||||||||||||||    || ||||||| |||||||||||||| || ||||||||
Db   324 GCTTCAGAAAATGAAGAGACACCAAAAGACCAGGAAGGATTAACT-----AATTTTTATGCTCAGATTT 270

Qy  1366 GATGAAGAAATGGATAGC-ATGTAGCAGTAGATTAGACATAAAAGAGATATGATCTTAGTTGGT 1424
         ||||  ||||  |||||| |||||| || |||||| | ||||||||||| ||||||| ||||||
Db   269 AATGCAGAAATAGTCAGCTATATCAAAGGTGATTGGATCAAAAATGGGTATGGATCTAATCGT 210

Qy  1425 AAAGGACTCAGGGGGAAAC---ACAGCAAGGCCCTGTCTATTCAGATTCTTTCTTGATCTCT 1481
         |||||||||||||||||||   ||||||||||| ||  |||| |||| ||| |||| |||||
Db   209 TATAGACTGGGGGGTGAGCGGACCAAGCAAGCCCTGTCTGTTCAGATTCTTTCTTGGACGCT 150
```

RESULT 11

```
Qy    1482 CTCTCTCTATGTATAGCAATTCTTCCTTCCTGAGTAGTGGGGCAGGACTCTTCTTCAATGAG 1541
Db     149 C------T------TGCAATGATTATTATTCCCTCTTGGGTATGGGGCCGGGACTTCTCTAAATGGAG 98
```

```
LOCUS       BG699761           777 bp    mRNA    linear    EST 07-MAY-2001
DEFINITION  602681454F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4814340 5',
            mRNA sequence.
ACCESSION   BG699761
VERSION     BG699761.1  GI:13968395
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10709  row: k  column: 13
            High quality sequence stop: 772.
FEATURES             Location/Qualifiers
     source          1..777
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4814340"
                     /tissue_type="hippocampus"
                     /clone_lib="NIH_MGC_95"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (grcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."
ORIGIN
```

```
Query Match     2.1%;  Score 172.4;  DB 12;  Length 777;
Best Local Similarity    99.4%;   Pred. No. 1e-23;
Matches   173;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps    0;
```

```
Qy  6315 AGCAGGAAGACGCTCAGAGAGGAGGAATTCTGACAATATCTTTACCGGAGGAGGAGGAGTAGC 6374
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 AGCGGGAAGACGCTCAGAGAGGAGGAATTCTGACAATATCTTTACCGGAGGAGGAGGAGTAGC 60

Qy  6375 TCAAAGCCGAAGCCACAGCTCCTCCTGCCGCGCCACATTCTTTCCTGCGTTGCGAATTCCAAGCT 6434
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCAAAGCCGAAGCCACCAGCTCCTCCTGCCGCGCCATTCTTTTCTTTCCTCGCTTGCGAAATTCCAAGCT 120
```

RESULT 12
BG699541
LOCUS   BG699541   593 bp   mRNA   linear   EST 07-MAY-2001
DEFINITION   602679246F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812059 5',
              mRNA sequence.
ACCESSION   BG699541
VERSION   BG699541.1 GI:13967945
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 593)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
              cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
              Toshiyuki and Piero Carninci (RIKEN)
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM10703 row: 1 column: 12
              High quality sequence stop: 591.
FEATURES   Location/Qualifiers
     source   1..593
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4812059"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."

ORIGIN

Query Match   2.1%;   Score 170; DB 12;   Length 593;
Best Local Similarity   100.0%;   Pred. No. 3.2e-23;
Matches 170;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   6435 GTTAAATAAGATGTGCAAAAGGGCTTGCAGGTCTCGCGGCGTTCTTGCTTGAGGAG 6488
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTTAAATAAGATGTGCAAAAGGGCTTGCAGGTCTCGCGGCGTTCTTGCTTGAGGAG 174

Qy   6435 GTTAAATAAGATGTGCAAAAGGGCTTGCAGGTCTCGCGGCGTTCTTGCTTGAGGAG 6488

Qy   6319 GGAAGACGCTTCAGAGGATTCTGACAACAATATCTTTACCGGAGAAGAGGCAAAGTACGGCTCAA 6378
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GGAAGACGCTTCAGAGGATTCTGACAACAATATCTTTACCGGAGAAGAGGCAAAGTACGGCTCAA 63

Qy   6379 AGCCGGAAGCCACAGCTCCTCCTCCGCCGCCATTTCTTCTTCCGCCGCCAATTCCAAGCTGTTA 6438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 AGCCGGAAGCCACAGCTCCTCCTCCGCCGCCATTTCTTCTTCCGCCGCCAATTCCAAGCTGTTA 123

Qy   6439 AATAAGATGTGCAAAAGGGCTTGCAGGTCTCGCGGCGTTCTTGCTTGAGGAG 6488
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 AATAAGATGTGCAAAAGGGCTTGCAGGTCTCGCGGCGTTCTTGCTTGAGGAG 173

RESULT 13

BI548918
LOCUS   BI548918   624 bp   mRNA   linear   EST 05-SEP-2001
DEFINITION   6031890043F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5260403 5',
              mRNA sequence.
ACCESSION   BI548918
VERSION   BI548918.1 GI:15436217
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 624)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
              cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
              Toshiyuki and Piero Carninci (RIKEN)
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM11656 row: e column: 12
              High quality sequence stop: 624.
FEATURES   Location/Qualifiers
     source   1..624
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5260403"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."

ORIGIN

Query Match   2.1%;   Score 170; DB 12;   Length 624;
Best Local Similarity   100.0%;   Pred. No. 3.2e-23;
Matches 170;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   6319 GGAAGACGCTCAGAGGATTCTGACAACAATATCTTTACCGGAGAAGAGGCAAAGTACGGCTCAA 6378
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GGAAGACGCTCAGAGGATTCTGACAACAATATCTTTACCGGAGAAGAGGCAAAGTACGGCTCAA 63

Qy   6379 AGCCGGAAGCCACAGCTCCTCCTCCGCCGCCATTTCTTTTCCTGCGCTTGCGGAATTCCAAGCTGTTA 6438
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 AGCCGGAAGCCACAGCTCCTCCTCCGCCGCCATTTCTTTTCCTGCGCTTGCGGAATTCCAAGCTGTTA 123

Qy   6439 AATAAGATGTGCAAAAGGGCTTGCAGGTCTCGCGGCGTTCTTGCTTGAGGAG 6488
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 AATAAGATGTGCAAAAGGGCTTGCAGGTCTCGCGGCGTTCTTGCTTGAGGAG 173

RESULT 14
BI548997
LOCUS   BI548997   680 bp   mRNA   linear   EST 05-SEP-2001
DEFINITION   6031889958F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5260358 5',
              mRNA sequence.
ACCESSION   BI548997
VERSION   BI548997.1 GI:15436309
KEYWORDS   EST.

SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 600)
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11656 row: c column: 15
            High quality sequence stop: 680.

FEATURES   Location/Qualifiers
     source   1..680
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5260358"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."

ORIGIN
Query Match   2.1%; Score 170; DB 12; Length 680;
Best Local Similarity 100.0%; Pred. No. 3.1e-23;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   6319 GGAAGACGCTCAGAGGAGGAGGATTCTGACAAATATCTTTACCGGAGAGGAGGCAAAGTACGCTCAA 6378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GGAAGACGCTCAGAGGAGGAGGATTCTGACAAATATCTTTACCGGAGAGGAGGCAAAGTACGCTCAA 63

Qy   6379 AGCCGAAGCCACAGCTCCTCCTGCCGCGCATTCTTTCTTCCTGCCGAATTCCAAGCTCGTTA 6438
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 AGCCGAAGCCACAGCTCCTCCTGCCGCGCATTCTTTCTTCCTGCCGAATTCCAAGCTCGTTA 123

Qy   6439 AATAAGAGTGTGCAAAGGGCTTGCAGGTCTGCCGCGGCTTCTTGCTTGAGGAG 6488
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 AATAAGAGTGTGCAAAGGGCTTGCAGGTCTGCCGCGGCTTCTTGCTTGAGGAG 173

RESULT 15
BI545493
LOCUS   BI545493   720 bp   mRNA   linear   EST 05-SEP-2001
DEFINITION 603187655F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5259084 5',
            mRNA sequence.
ACCESSION   BI545493
VERSION   BI545493.1 GI:15432805
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 720)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11652 row: n column: 13
            High quality sequence stop: 714.

FEATURES   Location/Qualifiers
     source   1..720
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5259084"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."

ORIGIN
Query Match   2.1%; Score 170; DB 12; Length 720;
Best Local Similarity 100.0%; Pred. No. 3.1e-23;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   6319 GGAAGACGCTCAGAGGAGGAGGATTCTGACAAATATCTTTACCGGAGAGGAGGCAAAGTACGCTCAA 6378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 GGAAGACGCTCAGAGGAGGAGGATTCTGACAAATATCTTTACCGGAGAGGAGGCAAAGTACGCTCAA 63

Qy   6379 AGCCGAAGCCACAGCTCCTCCTGCCGCGCATTCTTTCTTCCTGCCGAATTCCAAGCTCGTTA 6438
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 AGCCGAAGCCACAGCTCCTCCTGCCGCGCATTCTTTCTTCCTGCCGAATTCCAAGCTCGTTA 123

Qy   6439 AATAAGAGTGTGCAAAGGGCTTGCAGGTCTGCCGCGGCTTCTTGCTTGAGGAG 6488
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 AATAAGAGTGTGCAAAGGGCTTGCAGGTCTGCCGCGGCTTCTTGCTTGAGGAG 173

Search completed: September 11, 2004, 20:08:10
Job time : 12103 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: September 11, 2004, 08:15:38 ; Search time 19840 Seconds
                                                 (without alignments)
                                                 17566.600 Million cell updates/sec

Title:
Perfect Score: 8041
Sequence: 1 caaattgaagtcctggcac.............ctaggatctcaggtggatcc 8041
Scoring table: IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 2167151699S residues
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                   Maximum Match 100%
                   Listing first 45 summaries

Database :   GenEmbl:*
                   1: gb_ba:*
                   2: gb_htg:*
                   3: gb_in:*
                   4: gb_om:*
                   5: gb_ov:*
                   6: gb_pat:*
                   7: gb_ph:*
                   8: gb_pl:*
                   9: gb_pr:*
                   10: gb_ro:*
                   11: gb_sts:*
                   12: gb_sy:*
                   13: gb_un:*
                   14: gb_vi:*
                   15: gb_ba:*
                   16: em_fun:*
                   17: em_hum:*
                   18: em_in:*
                   19: em_mu:*
                   20: em_om:*
                   21: em_or:*
                   22: em_ov:*
                   23: em_pat:*
                   24: em_ph:*
                   25: em_pl:*
                   26: em_ro:*
                   27: em_sts:*
                   28: em_un:*
                   29: em_vi:*
                   30: em_htg_hum:*
                   31: em_htg_inv:*
                   32: em_htg_mus:*
                   33: em_htg_other:*
                   34: em_htg_pln:*
                   35: em_htg_rod:*
                   36: em_htg_mam:*
                   37: em_htg_vrt:*
                   38: em_sy:*
                   39: em_htgo_hum:*
                   40: em_htgo_mus:*
                   41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8041	100.0	20300	6	AX451337	AX451337 Sequence
2	8039.4	100.0	165329	9	AL583850	AL583850 Human DNA
3	8009.2	99.6	191699	9	AC031977	AC031977 Homo sapi
4	426	5.3	275631	2	AC125563	AC125563 Rattus no
5	402.6	5.0	224486	2	AC115766	AC115766 Mus muscu
6	326.2	4.1	172931	2	AC006581	AC006581 Homo sapi
7	317	3.9	94158	2	AC119619	AC119619 Homo sapi
8	317	3.9	154125	2	AC137626	AC137626 Homo sapi
9	311.2	3.9	154125	2	AC073227	AC073227 Homo sapi
10	311.2	3.9	196832	2	AC018694	AC018694 Homo sapi
11	307.2	3.8	176773	2	AC009867	AC009867 Homo sapi
12	305.6	3.8	89000	2	AP003059	AP003059 Homo sapi
13	299.4	3.7	169972	9	AL513284	AL513284 Human DNA
14	297.6	3.7	121720	9	AL591491	AL591491 Human DNA
15	293.8	3.7	134937	9	AC010244	AC010244 Homo sapi
16	293.8	3.7	170219	9	AC091991	AC091991 Homo sapi
17	293.8	3.7	182049	2	AC021239	AC021239 Homo sapi
18	293.8	3.7	188207	2	AC021800	AC021800 Homo sapi
19	292	3.6	193126	9	AL139275	AL139275 Human DNA
20	290.4	3.6	158983	2	AC073070	AC073070 Homo sapi
21	289.6	3.6	150145	9	AC069216	AC069216 Homo sapi
22	289.6	3.6	178168	9	AC024412	AC024412 Homo sapi
23	289	3.6	126394	9	AC109998	AC109998 Homo sapi
24	289	3.6	194860	9	AC091903	AC091903 Homo sapi
25	288.2	3.6	65608	9	AL450243	AL450243 Human DNA
26	288.2	3.6	181618	9	AL442129	AL442129 Homo sapi
27	288.2	3.6	190092	2	AC120502	AC120502 Lemur cat
28	284.4	3.5	147055	9	AL023580	AL023580 Human DNA
29	282.2	3.5	178905	9	HS372K1	HS372K1 Human DNA
30	280.2	3.5	98232	9	AC024153	AC024153 Homo sapi
31	280.2	3.5	152005	9	AC012271	AC012271 Homo sapi
32	280	3.5	27724	9	AC100834	AC100834 Homo sapi
33	280	3.5	104011	9	BX284678	BX284678 Human DNA
34	276.2	3.4	137996	9	AL590392	AL590392 Human DNA
35	275.8	3.4	160800	2	AP001982	AP001982 Homo sapi
36	275.8	3.4	162584	2	AP002011	AP002011 Homo sapi
37	275.8	3.4	183674	2	AP002763	AP002763 Homo sapi
38	275.4	3.4	142533	9	BS000038	BS000038 Pan trogl
39	274	3.4	228353	9	AC016996	AC016996 Homo sapi
40	273.8	3.4	172461	2	AL445283	AL445283 Homo sapi
41	273.8	3.4	173010	9	AC083796	AC083796 Homo sapi
42	271.6	3.4	154336	9	CN501DSU	CN501DSU Human chr
43	271.6	3.4	154341	9	HS1164K7	HS1164K7 Homo sapi
44	268.8	3.3	204662	9	CNS01RHQ	CNS01RHQ Human chr
45	267.2	3.3	127790	9	AC008885	AC008885 Homo sapi

ALIGNMENTS

RESULT 1
AX451337
LOCUS   AX451337   20300 bp   DNA   linear   PAT 03-JUL-2002
DEFINITION   Sequence 3 from Patent WO0216653.
ACCESSION   AX451337
VERSION   AX451337.1   GI:21698388
KEYWORDS   .
SOURCE   synthetic construct
  ORGANISM   synthetic construct
                artificial sequences.
REFERENCE   1
  AUTHORS   Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
  TITLE   Methods and systems for facilitating the diagnosis and treatment of
                schizophrenia.
  JOURNAL   Patent: WO 0216653-A 3 28-FEB-2002;

University of Pittsburgh (US)

FEATURES   Location/Qualifiers
source   1..20300
    /organism="synthetic construct"
    /mol_type="unassigned DNA"
    /db_xref="taxon:32630"
    /note="A genomic sequence containing RGS4 nucleic acid
    sequence and sequences upstream and downstream to the
    RGS4 nucleic acid sequence"

ORIGIN

Query Match   100.0%;   Score 8041;   DB 6;   Length 20300;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 8041;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 CAAATTGAAGTCCTGGGCCACCTTACAAAACTAGAGGTCATAAAGTTTACCCTTAATTACC   60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4110 CAAATTGAAGTCCTGGGCCACCTTACAAAACTAGAGGTCATAAAGTTTACCCTTAATTACC   4169

Qy   61 AAGATTTCCTAGAGAGATCCTGAGGATCTATAATTGGAATCCAGAGTTCTTGTAAGTTCAAGCA   120
Db   4170 AAGATTTCCTAGAGAGATCCTGAGGATCTATAATTGGAATCCAGAGTTCTTGTAAGTTCAAGCA   4229

Qy   121 CTTTCCATGACACCATACTCGTTTCTTTCCACCTGCACAATGCAAATGAACTCTTATGAAA   180
Db   4230 CTTTCCATGACACCATACTCGTTTCTTTCCACCTGCACAATGCAAATGAACTCTTATGAAA   4289

Qy   181 CTGCTGTTTCTATCCTGGGGCTAAATGTTGGCAGAAAAAAGATTTAATCTTTGGGATAAGC   240
Db   4290 CTGCTGTTTCTATCCTGGGGCTAAATGTTGGCAGAAAAAAGATTTAATCTTTGGGATAAGC   4349

Qy   241 TATTTTGGGGTTTTCTCCTACTTCTTGTTGGGAAAACAAAGGTTTTCTTCTCCTGGCTAATTAAGT   300
Db   4350 TATTTTGGGGTTTTCTCCTACTTCTTGTTGGGAAAACAAAGGTTTTCTTCTCCTGGCTAATTAAGT   4409

Qy   301 GTGGTATTTGTTCTTCCAGGGAAATCAGTGAATGCATCACCTGCTGCTAATCACGTGCGTGCTAATCAGGG   360
Db   4410 GTGGTATTTGTTCTTCCAGGGAAATCAGTGAATGCATCACCTGCTGCTAATCACGTGCGTGCTAATCAGGG   4469

Qy   361 TTGGAGTTCCTGATTTATTGTTCCATGTGCCCACAAAAGCTTGGTCACACTGTGCAAATGGACACATT   420
Db   4470 TTGGAGTTCCTGATTTATTGTTCCATGTGCCCACAAAAGCTTGGTCACACTGTGCAAATGGACACATT   4529

Qy   421 TCCCAAAAGTAAGACATACTGGGGAAGTCCCGTTGTTACCTTCCTGGTATACAGCATCCTCC   480
Db   4530 TCCCAAAAGTAAGACATACTGGGGAAGTCCCGTTGTTACCTTCCTGGTATACAGCATCCTCC   4589

Qy   481 AGCCCCATATCTTTGCTTTTAGTCCTAAAAATCAATAACTGAACCTCTCATTGATGTGTCTA   540
Db   4590 AGCCCCATATCTTTGCTTTTAGTCCTAAAAATCAATAACTGAACCTCTCATTGATGTGTCTA   4649

Qy   541 GGCCAGTTGTAGTAAACAAATAAAAGAAGAGGGGGGAGGCGGCTTGACAAACTGAGAGAAAATTGTC   600
Db   4650 GGCCAGTTGTAGTAAACAAATAAAAGAAGAGGGGGGAGGCGGCTTGACAAACTGAGAGAAAATTGTC   4709

Qy   601 ATCTGAAGTGGTGCAAAGCACAGCCTGGGGGCTGAGCCCTTGGCCTCCTACATCCTGCCCAAGTGG   660
Db   4710 ATCTGAAGTGGTGCAAAGCACAGCCTGGGGGCTGAGCCCTTGGCCTCCTACATCCTGCCCAAGTGG   4769

Qy   661 AGGATCAGTGCCCCATTTAACATCTGGTGAGAACTAAAGAACGGCAAAGGCTGCCACAATGA   720
Db   4770 AGGATCAGTGCCCCATTTAACATCTGGTGAGAACTAAAGAACGGCAAAGGCTGCCACAATGA   4829

Qy   721 CTTATTTCCCTGCATTTGATGATCCGGTCAATCCTTGAGAAAATGTTTTCTTTTGTTCTCCCTG   780
Db   4830 CTTATTTCCCTGCATTTGATGATCCGGTCAATCCTTGAGAAAATGTTTTCTTTTGTTCTCCCTG   4889

Qy   781 AGCAAAGGTTGGAAAAATTTGAAATTTACCTAGAGACCACACATAGTTCACATCCTGCCTG   840
Db   4890 AGCAAAGGTTGGAAAAATTTGAAATTTACCTAGAGACCACACATAGTTCACATCCTGCCTG   4949

Qy   841 TGTGGCTGAATGTCTGCCCCCAGTAGAGAAAACAGTTCTTCTAAAGCCTTATTGTCAACAAT   900
Db   4950 TGTGGCTGAATGTCTGCCCCCAGTAGAGAAAACAGTTCTTCTAAAGCCTTATTGTCAACAAT   5009

Db   4950 TGTGGCTGAATGTCTGCCCCCAGTAGAGAAAACAGTTCTTCTAAAGCCTTATTGTCAACAAT   5009

Qy   901 ACCTTCCAGATGTTGCATTTACAAATTTAAGAAATTAAGAAATAGCTTCAAACTTTTTG   960
Db   5010 ACCTTCCAGATGTTAGCATTTTACAAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTG   5069

Qy   961 CCAGTTTCTCTGATAATCAAATCTATTCTTTTACTACTGCCTCCCAAGCTTCCTAGAA   1020
Db   5070 CCAGTTTCTCTGATAATCAAATCTATTCTTTTACTACTGCCTCCCAAGCTTCCTAGAA   5129

Qy   1021 TGCTAACCTGATCGGCTTAAGTACTTGAACTACCTCTTCCTCCTCCCATTAACTACACAGTA   1080
Db   5130 TGCTAACCTGATCGGCTTAAGTACTTGAACTACCTCTTCCTCCTCCCATTAACTACACAGTA   5189

Qy   1081 AATTCTGGGTCTTCAGCAGTAACAACAGAAAAAACACCCTTTAGTTCTCAGCATATTCGGTGCACCTT   1140
Db   5190 AATTCTGGGTCTTCAGCAGTAACAACAGAAAAAACACCCTTTAGTTCTCAGCATATTCGGTGCACCTT   5249

Qy   1141 CATTTATCTCTCCTTCTCTCCTCAAAGCTGCAGTAGGGTGAAAACGTGTGATACAGTTTTC   1200
Db   5250 CATTTATCTCTCCTTCTCTCCTCAAAGCTGCAGTAGGGTGAAAACGTGTGATACAGTTTTC   5309

Qy   1201 TCTTCCATCATAAGGGTGGCACCAAAACTCCTATAGTAAGAGACAGGTTAATAGAGGCA   1260
Db   5310 TCTTCCATCATAAGGGTCGCACCAAAACTCCTATAGTAAGAGACAGGTTAATAGAGGCA   5369

Qy   1261 AAAACCTAACAAAATTTATTTAATCAAAGTTTACATGACATGGGAGTGTTTCAGAAATGAAG   1320
Db   5370 AAAACCTAACAAAATTTATTTAATCAAAGTTTACATGACATGGGAGTGTTTCAGAAATGAAG   5429

Qy   1321 ACCCAAAAGACCCAGGGGGAAAACTGTCTGTTTTTTTTTTGGCTGAGGGTTCGATGAAGAATGGATA   1380
Db   5430 ACCCAAAAGACCCAGGGGGAAAACTGTCTGTTTTTTTTTTGGCTGAGGGTTCGATGAAGAATGGATA   5489

Qy   1381 GCAGTAGCCATGTAGATTAGACAAAAAGGATAAGATGATTCAGATTCTTAGTGGACTCAGGGGGA   1440
Db   5490 GCAGTAGCCATGTAGATTAGACAAAAAGGATAAGATGATTCAGATTCTTAGTGGACTCAGGGGGA   5549

Qy   1441 AACACAGCAAGGCTGTCATTCAGATTCTGATCTCTCTCTCTAGTATAATAGCAT   1500
Db   5550 AACACAGCAAGGCTGTCATTCAGATTCTGATCTCTCTCTCTAGTATAATAGCAT   5609

Qy   1501 TCTTTCCTCCTGAGTATGGGGTAGGGCCAGGACTCCTTCTTCTTCTTGGGGTCTTCTAGCAT   1560
Db   5610 TCTTTCCTCCTGAGTATGGGGTAGGGCCAGGACTCCTTCTTCTTCTTGGGGTCTTCTAGCAT   5669

Qy   1561 AGAAAGTGGCCCTTTTAGATTTAGGCTTGGTTCGGGGAGAGGAGTTCTAGTTTCTAT   1620
Db   5670 AGAAAGTGGCCCTTTTAGATTTAGGCTTGGTTCGGGGAGAGGAGTTCTAGTTTCTAT   5729

Qy   1621 GACCCATCTTGGGGAGAGGAATTCTGGTTTCTGGCTTCATGAAGAAAAGAGGA   1680
Db   5730 GACCCATCTTGGGGAGAGGAATTCTGGTTTCTGGCTTCATGAAGAAAAGAGGA   5789

Qy   1681 GTAAGAGGCAGGAGGGCAGGAGGTCAGAAAGAGAGACTTGGCTTGCTTCC   1740
Db   5790 GTAAGAGGCAGGAGGGCAGGAGGTCAGAAAGAGAGACTTGGCTTGCTTCC   5849

Qy   1741 GCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTACTACTTTGGCATATGGT   1800
Db   5850 GCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTACTACTTTGGCATATGGT   5909

Qy   1801 TTTCTGAGCTCTAACACTGCAATCATGCTAAACTCCTCTATGACCTTCAAACATTCACT   1860
Db   5910 TTTCTGAGCTCTAACACTGCAATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACT   5969

Qy   1861 TGCTTTTATTCTTTATGGTTGTGATGGCATAGAGGTCAATAGCAAAGACAAAGAAGAGGTA   1920
Db   5970 TGCTTTTATTCTTTATGGTTGTGATGGCATAGAGGTCAATAGCAAAGACAAAGAAGAGGTA   6029

Qy   1921 ACTGTCTGAGCTGGCATAACATTACTACCACTTAATCAAATGTAAGTGTAAGCTCAGGTAAGTAC   1980
Db   6030 ACTGTCTGAGCTGGCATAACATTACTACCACTTAATCAAATGTAAGTGTAAGCTCAGGTAAGTAC   6089

```
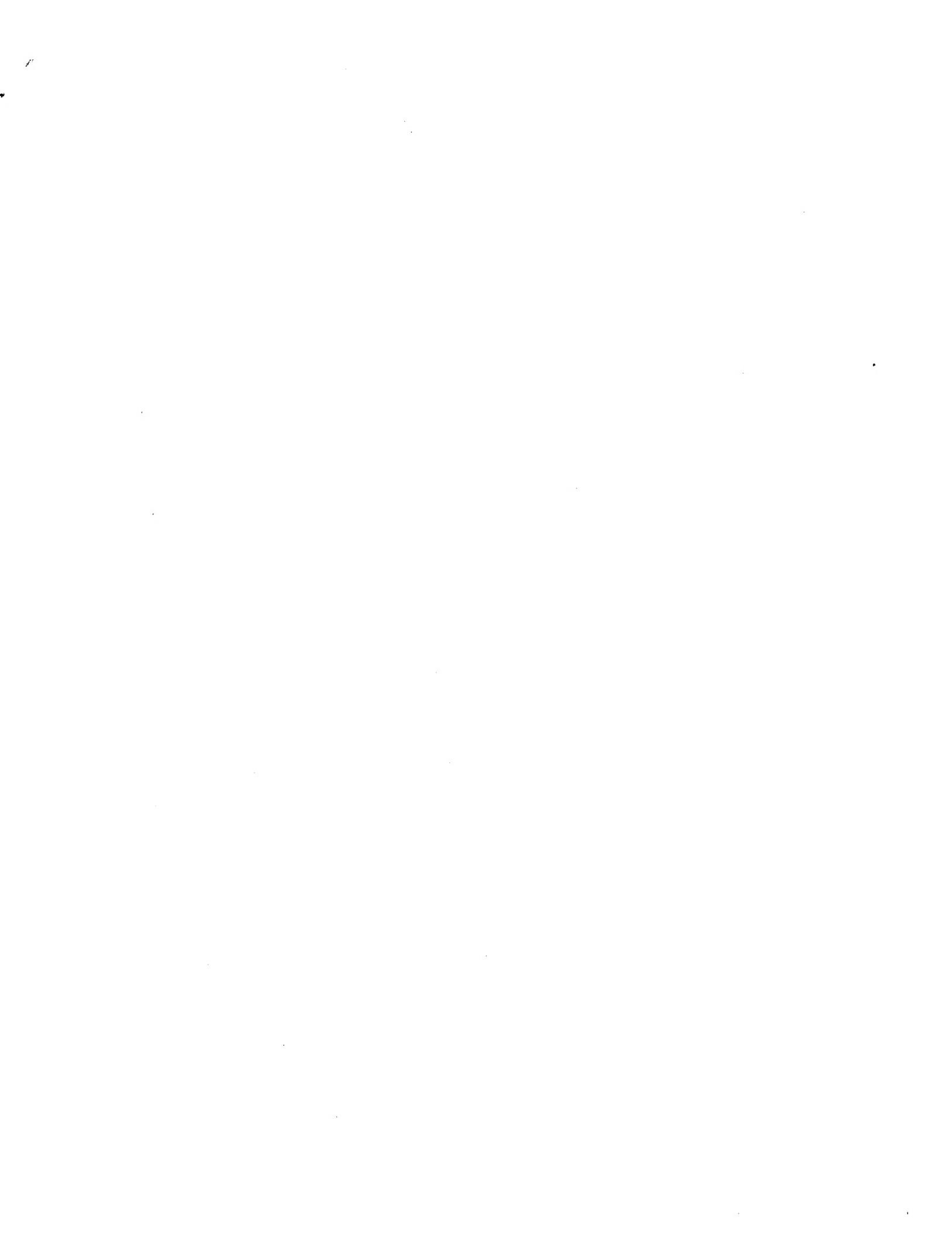Qy   6361 GAGGCAAAGTACGGTCAAAGCCGAAGCCACAGCTCCTCCTGCCGCCGCGCATTCTTTCTTCCTGCCTT 6420
Db  10470 GAGGCAAAGTACGGTCAAAGCCGAAGCCACAGCTCCTCCTGCCGCCGCGCATTCTTTCTTCCTGCCTT 10529

Qy   6421 GCGAATTCCAAGCTCTTAAATAAGATGTCAAAGGCTTGCAGGTCTGCCGCCTTCTTGC 6480
Db  10530 GCGAATTCCAAGCTCTTAAATAAGATGTCAAAGGCTTGCAGGTCTGCCGCCTTCTTGC 10589

Qy   6481 TTGAGGAGGTAAAGATTGTTTCAGCCATTAAACTTTTTGGCTAGACTTTTCTC 6540
Db  10590 TTGAGGAGGTAAAGATTGTTTCAGCCATTAAACTTTTTGGCTAGACTTTTCTC 10649

Qy   6541 AGTTATTTACATGTTGTACTTACTTACATATATAAAACTTGTGGTCAG 6600
Db  10650 AGTTATTTACATGTTGTACTTACTTACATATATAAAACTTGTGGTCAG 10709

Qy   6601 GAGAGACACGACTTTCTAACTTCCTCCAAGACTAGACATGT 6660
Db  10710 GAGAGACACGACTTTCTAACTTCCTCCAAGACTAGACATGT 10769

Qy   6661 GCTCCCCAAATTTCAGCGCCCTTATGTGTTTGAGGAACTGTTC 6720
Db  10770 GCTCCCCAAATTTCAGCGCCCTTATGTGTTTGAGGAACTGTTC 10829

Qy   6721 TATTCTTTTAAGCCAGGTCTAAGAAAGCTAGTTTTAATTAAGAAGCAGAGTGTTTC 6780
Db  10830 TATTCTTTTAAGCCAGGTCTAAGAAAGCTAGTTTTAATTAAGAAGCAGAGTGTTTC 10889

Qy   6781 GCTATGTACAGTGATCTGTAATAATCTCCATCTGTGATTAGTTACTATCCC 6840
Db  10890 GCTATGTACAGTGATCTGTAATAATCTCCATCTGTGATTAGTTACTATCCC 10949

Qy   6841 TGGAGTACATAGAAGCCTGGCTCGGGCTTCGAATTGTAGTACAACTTGTTTCAGGA 6900
Db  10950 TGGAGTACATAGAAGCCTGGCTCGGGCTTCGAATTGTAGTACAACTTGTTTCAGGA 11009

Qy   6901 AAGGTACCCCAGAGAATGAGGTTTGGCTCCATCAGAAAGGCACTATGCTTCGTGTGG 6960
Db  11010 AAGGTACCCCAGAGAATGAGGTTTGGCTCCATCAGAAAGGCACTATGCTTCGTGTGG 11069

Qy   6961 TGGTGCAGTAACTTTCACTCTCTARTGTTCTTATAAGCAAAATGTTACAATGAGATATGAGT 7020
Db  11070 TGGTGCAGTAACTTTCACTCTCTARTGTTCTTATAAGCAAAATGTTACAATGAGATATGAGT 11129

Qy   7021 TTTAAAGCCAGATCTTCCTTATCTCTCGCCCCATCTCTTCTTGAAGTGTCTCATAT 7080
Db  11130 TTTAAAGCCAGATCTTCCTTATCTCTCGCCCCATCTCTTCTTGAAGTGTCTCATAT 11189

Qy   7081 GAGTTTGGTTGGAGAAATATTGATCATTACAAAATCAGTTAATAGTTTTGTTAGAAGATCTCA 7140
Db  11190 GAGTTTGGTTGGAGAAATATTGATCATTACAAAATCAGTTAATAGTTTTGTTAGAAGATCTCCA 11249

Qy   7141 TCTTAAAGACATTGTTTGTTTTTGTTAATAATACTCCCTTGATTTTTTAAAAAGACCTTACAGACA 7200
Db  11250 TCTTAAAGACATTGTTTTGTTTTTGTTAATAATACTCCCTTGATTTTTTAAAAAGACCTTACAGACA 11309

Qy   7201 TACAGCCTATTCATTTGTTTTGTTTGGTTTTGGTTGTTCAAATTGCATTCAGGAGA 7260
Db  11310 TACAGCCTATTCATTTGTTTTGTTTGGTTTTGGTTGTTCAAATTGCATTCAGGAGA 11369

Qy   7261 AAGATCATATATTAGCCAGTTGAAAATTTGAAAATTAAAACAAAATGAGTGGTATATAA 7320
Db  11370 AAGATCATATATTAGCCAGTTGAAAATTTGAAAATTAAAACAAAATGAGTGGTATATAA 11429

Qy   7321 ATCTTGCAGTCAAAAGGTAAAAAGTAGAAGTCAACCTAAAAGGTATATCTTAATCCGAC 7380
Db  11430 ATCTTGCAGTCAAAAGGTAAAAAGTAGAAGTCAACCTAAAAGGTATATCTTAATCCGAC 11489

Qy   7381 TGCAAATAGAAATTACCACAAAATTTATTTTTGGAAATAATCTCAGAAAACATAATTTTT 7440
Db  11490 TGCAAATAGAAATTACCACAAAATTTATTTTTGGAAATAATCTCAGAAAACATAATTTTT 11549
```

```
Qy   7441 ATGTACTATTAAAACATTTACTTTTCCAAATATTCTGTCATTCAGGAGGTATGGAAGTATCG 7500
Db  11550 ATGTACTATTAAAACATTTACTTTTCCAAATATTCTGTCATTCAGGAGGTATGGAAGTATCG 11609

Qy   7501 ATGGGCTTCTTTAAAATGAAGCACAGGAGGCGTCTTGGCAGGAGAGTAGTTATCTATTGAAATAAGTTCCT 7560
Db  11610 ATGGGCTTCTTTAAAATGAAGCACAGGAGGCGTCTTGGCAGGAGAGTAGTTATCTATTGAAATAAGTTCCT 11669

Qy   7561 CTGACCTTCACGCTTAATTCTTGAAATGGAGGAGAGTGGAGCAAAATTACTTCACTTA 7620
Db  11670 CTGACCTTCACGCTTAATTCTTGAAATGGAGGAGAGTGGAGCAAAATTACTTCACTTA 11729

Qy   7621 CTTGCATATGAAAATGAAACCGTACAAAAATTACAAGAAGGTCAGGAGAAAAGTTATGCTCTGG 7680
Db  11730 CTTGCATATGAAAATGAAACCGTACAAAAATTACAAGAAGGTCAGGAGAAAAGTTATGCTCTGG 11789

Qy   7681 TAAATATTTTTGCAAAACAGATAAAAGAATAAATACTAGAGACTCTTCCTCCAAAGGAGTTAAGC 7740
Db  11790 TAAATATTTTTGCAAAACAGATAAAAGAATAAATACTAGAGACTCTTCCTCCAAAGGAGTTAAGC 11849

Qy   7741 AGCTAAATCTAAGGAGAGTGAAACTTATGTCAGCAGAAGATGAAACTGTCTCTTCCCTTTCTTCCT 7800
Db  11850 AGCTAAATCTAAGGAGAGTGAAACTTATGTCAGCAGAAGATGAAACTGTCTCTTCCCTTTCTTCCT 11909

Qy   7801 CAATAAATTGCAAATCAATCCTAGTCCAACATCATCTTTACCACCAGTGGCCTCAGACAAAATCTAGAAATC 7860
Db  11910 CAATAAATTGCAAATCAATCCTAGTCCAACATCATCTTTACCACCAGTGGCCTCAGACAAAATCTAGAAATC 11969

Qy   7861 AGCCAATTGCCTTCTCAAGGTCACATAGGTGGTCGTGGGGTGGGTTAGGACCAAAATCTAGAATTC 7920
Db  11970 AGCCAATTGCCTTCTCAAGGTCACATAGGTGGTCGTGGGGTGGGTTAGGACCAAAATCTAGAATTC 12029

Qy   7921 CTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTGTTTTATTTTTATGGTTTTATTATAAGGAA 7980
Db  12030 CTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTGTTTTATTTTTATGGTTTTATTATAAGGAA 12089

Qy   7981 TACTTGCTTAAGCACACTTACCCCCTGCATTGATTAATAACTTCTAGGATCTCAGGTGGATC 8040
Db  12090 TACTTGCTTAAGCACACTTACCCCCTGCATTGATTAATAACTTCTAGGATCTCAGGTGGATC 12149

Qy   8041 C 8041
Db  12150 C 12150
```

```

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-430G6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-430G6 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-430G6 is at 165329 in this sequence. The true right end of clone RP11-331H2 is at 2000 in this sequence.

FEATURES             Location/Qualifiers
     source          1..165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"

ORIGIN

Query Match             100.0%;  Score 8039.4;  DB 9;  Length 165329;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 8040;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

```
Qy   1 CAAAATTGAAGTCCCTGGGCCACGTTACAAAACTAGAGGTCATTAAAGTTTTACCCTAATTTACC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75791 CAAAATTGAAGTCCCTGGGCCACGTTACAAAACTAGAGGTCATTAAAGTTTTACCCTAATTTACC 75850

Qy   61 AAGATTTCCTAGAGGATCATAAATTGGAATCCAGAATCTGCCCTTCTCTGTAAAGTTCAAGCA 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75851 AAGATTTCCTAGAGGATCATAAATTGGAATCCAGAATCTGCCCTTCTCTGTAAAGTTCAAGCA 75910

Qy   121 CTTTCCATGACACCATACTGTTTCTTTCCACCTTGCACAAATGAAAGTAAACTCTTATGAAA 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75911 CTTTCCATGACACCATACTGTTTCTTTCCACCTTGCACAAATGAAAGTAAACTCTTATGAAA 75970

Qy   181 CTGCTGTTTCTATCCTCTGGGGTAAATGTTGCCAGAAAAAGAGATTTAAATCTTTGGGATAAGGC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75971 CTGCTGTTTCTATCCTCTGGGGTAAATGTTGCCAGAAAAAGAGATTTAAATCTTTGGGATAAGGC 76030

Qy   241 TATTTTGGGTTTTCTTCCTACTTCCTTGGGAAACAAGGTTTTCTTCCTCCCGGGCTAATTAAGT 300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76031 TATTTTGGGTTTTCTTCCTACTTCCTTGGGAAACAAGGTTTTCTTCCTCCCGGGCTAATTAAGT 76090

Qy   301 GTGGTATTTGTTCTTCCAGGGAAATCAGTGAATGCATCACCTGCTGTCAATCAAAGTCAGG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76091 GTGGTATTTGTTCTTCCAGGGAAATCAGTGAATGCATCACCTGCTGTCAATCAAAGTCAGG 76150

Qy   361 TTGGAGTTCCTGATTTATTGCATGTGCCCACAAAGCTTGGTGCAAAGAATTGGACACATT 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76151 TTGGAGTTCCTGATTTATTGCATGTGCCCACAAAGCTTGGTGCAAAGAATTGGACACATT 76210

Qy   421 TCCCAAAGTAAGACATACTGGGAAGTCCCTGTTTACCTTCCTGGTATACAGGCATCCTCTC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76211 TCCCAAAGTAAGACATACTGGGAAGTCCCTGTTTACCTTCCTGGTATACAGGCATCCTCTC 76270

Qy   481 AGCCCCATATCTTTGCTTTTTAGTCCTAAAAATCAATAACTTCATTGATGTCTA 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76271 AGCCCCATATCTTTGCTTTTTAGTCCTAAAAATCAATAACTTCATTGATGTCTA 76330

Qy   541 GGCCATTGTAGTAAACAATAAAGAAGGAGGAGGAGGCTTCTGAGACAAACTGAGAGAAAATTGTC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76331 GGCCATTGTAGTAAACAATAAAGAAGGAGGAGGAGGCTTCTGAGACAAACTGAGAGAAAATTGTC 76390

Qy   601 ATCTGAAGTGGTGCAAGCACAGCCTGGGGCTGAGCCTTGGCGTACATCCTGCCCAAGTGG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76391 ATCTGAAGTGGTGCAAGCACAGCCTGGGGCTGAGCCTTGGCGTACATCCTGCCCAAGTGG 76450

Qy   661 AGGATCAGGTGCCCCATTTAACATCTGGTAGAGAACTAAAGGAACGCGCTGCCACAATGA 720
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76451 AGGATCAGGTGCCCCATTTAACATCTGGTAGAGAACTAAAGGAACGCGCTGCCACAATGA 76510

Qy   721 CTTAATTCCCTGCATTTGATACCGTCAATCCTTGACGAATGTTTTCTTTTTTGTTCTCCCTG 780
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76511 CTTAATTCCCTGCATTTGATACCGTCAATCCTTGACGAATGTTTTCTTTTTTGTTCTCCCTG 76570

Qy   781 AGCAAAGGTGGGAAAAATTGAAAATTTACCTAGAGACCACACATAGTTTCACATCCTCTG 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76571 AGCAAAGGTGGGAAAAATTGAAAATTTACCTAGAGACCACACATAGTTTCACATCCTCTG 76630

Qy   841 TGTGGCTGAATGTTCTGCCCCCCAGTAGAAAACAGTTCTTCTTCAAGCCTTATGTCAACAAT 900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76631 TGTGGCTGAATGTTCTGCCCCCCAGTAGAAAACAGTTCTTCTTCAAGCCTTATGTCAACAAT 76690

Qy   901 ACCTTCCAGATGTTAGCATTTACAAATTAAGGAACTTAAAATAGCCTTCAAACTTTTG 960
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76691 ACCTTCCAGATGTTAGCATTTACAAATTAAGGAACTTAAAATAGCCTTCAAACTTTTG 76750

Qy   961 CCAGTTTCTCTGATATCCAATCTCATTCTTTTACTCTGCCTCCCAAGCTTCTTCTTTCTGAA 1020
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76751 CCAGTTTCTCTGATATCCAATCTCATTCTTTTACTCTGCCTCCCAAGCTTCTTCTTTCTGAA 76810

Qy   1021 TGCTAACCTGATCGGCTTAAGTACTTGAACTACCTTCTCTCCTCCATTAACTACAGAGTA 1080
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76811 TGCTAACCTGATCGGCTTAAGTACTTGAACTACCTTCTCTCCTCCATTAACTACAGAGTA 76870

Qy   1081 AATTCCTGGTCTTCAGAGTAACAAGAAAACACCCTTTAGTTCTTCAGCATTTCCTGCACCTT 1140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76871 AATTCCTGGTCTTCAGAGTAACAAGAAAACACCCTTTAGTTCTTCAGCATTTCCTGCACCTT 76930

Qy   1141 CATTTATCTCTCCTTCTCTTCTTTCTCAAAGTCTGCATTGAGGGTGAAAAGTTGTGATACATTTTC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76931 CATTTATCTCTCCTTCTCTTCTTTCTCAAAGTCTGCATTGAGGGTGAAAAGTTGTGATACATTTTC 76990

Qy   1201 TCTTCCATCATAAGGGTCGGCAACCAAAACAAAAACTCCTATAGTAATAAGAGAGCA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76991 TCTTCCATCATAAGGGTCGGCAACCAAAACAAAAACTCCTATAGTAATAAGAGAGCA 77050

Qy   1261 AAACCTAACAAAATTTATTTAATCAAAGTTTACATGACATGGGAGTCTTCAGAAATGAAG 1320
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77051 AAACCTAACAAAATTTATTTAATCAAAGTTTACATGACATGGGAGTCTTCAGAAATGAAG 77110

Qy   1321 ACCCAAAAGACCCAGGGGAAACTGTCTGTTGTTTTTTGCTGGTCATGAAGAATGGATA 1380
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77111 ACCCAAAAGACCCAGGGGAAACTGTCTGTTGTTTTTTGCTGGTCATGAAGAATGGATA 77170

Qy   1381 GCATGTAGCCATGTAGATTAGACAAAAAGGATGATCTTAGTGGTGAAGGAGGACTCAGGGGA 1440
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77171 GCATGTAGCCATGTAGATTAGACAAAAAGGATGATCTTAGTGGTGAAGGAGGACTCAGGGGA 77230

Qy   1441 AACACAGCAAAGGCCTGTCTTCATTCAGATTCTTTTGATCTTCCTCTCTCTCTTCTAGTGAGG 1500
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77231 AACACAGCAAAGGCCTGTCTTCATTCAGATTCTTTTGATCTTCCTCTCTCTCTTCTAGTGAGG 77290

Qy   1501 TCTTTCCCTCCTCTGACTATGGGCAGGACTCTCGTCAATGAGTAGTATAGCAT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77291 TCTTTCCCTCCTCTGACTATGGGCAGGACTCTCGTCAATGAGTAGTATAGCAT 77350

Qy   1561 AGAAAGTGGCCTTTTACAGATTTATGGCTTGCTTCGGGGAGGAGGAGGTTCTAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77351 AGAAAGTGGCCTTTTACAGATTTATGGCTTGCTTCGGGGAGGAGGAGGTTCTAT 77410

Qy   1621 GACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAGAAAGAGGA 1680
```

Qy   3841 TCCTCAAACACAATCTGCAACAGTGCCCACACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTT 3900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 79631 TCCTCAAACACAATCTGCAAGTCCCACAGTGAAAAAAGTAAGTTAACTCATGTGAAGCGTT 79690

Qy   3901 TTACAAACACTTTTTTAAAAGTCTTAAAAGAAAAGCAAGATTTAATAGTCAAAG 3960
           |||||||||||||||||||||||||||||||||||||||||||||
Db 79691 TTACAAACACTTTTTTAAAAGTCTTAAGAAAGCAAGATTTAATAGTCAAAG 79750

Qy   3961 AAGTGAGTAAACATGAAAATGGCCTGAACAGAGTAATGAGCTTAGAGACAT 4020
Db 79751 AAGTGAGTAAACATGAAAATGCCTGAACAGAGTAATGAGCTTAGAGACAT 79810

Qy   4021 GTTAGTTAATATGTCTTGAAAGCAGCAGCCTCCTGCTTCAAGGACAAATTGG 4080
Db 79811 GTTAGTTAATATGTCTTGAAAGCAGCAGCCTCCTGCTTCAAGGACAAATTGG 79870

Qy   4081 CAAGTGAACACTCCTTGAAATAAAAGTGTAAAAATAATTTGTTTTCTATACTGT 4140
Db 79871 CAAGTGAACACTCCTTGAAATAAAAGTGTAAAAATAATTTGTTTTCTATACTGT 79930

Qy   4141 GTATAATAGAATGATAAAAATTATTTTGACTAGACACTTTGTTAGAAAATATCTCTATT 4200
Db 79931 GTATAATAGAATGATAAAAATTATTTTGACTAGACACTTTGTTAGAAAATATCTCTATT 79990

Qy   4201 TACACAGTTTACCTTATTGATAAGACTGTTGAGTGAATGAAGAAGCTGTAA 4260
Db 79991 TACACAGTTTACCTTATTGATAAGACTGTTGAGTGAATGAAGAAGCTGTAA 80050

Qy   4261 ACATAACTGAGTATCGGAGACACCTGTAATCGGGAATAGCAGGTGGACAAATCC 4320
Db 80051 ACATAACTGAGTATCGGAGACACCTGTAATCGGGAATAGCAGGTGGACAAATCC 80110

Qy   4321 CCTCAGCGAAGTCACTTTCTCTCTTGGGTCTCTATTTCCTTTTGGGTGAAATGAGAGTGT 4380
Db 80111 CCTCAGCAAGTCACTTTCTCTTTCTGGGGTCTCTATTCCTTTTGGGTGAAATGAGAGTGT 80170

Qy   4381 TAGGCTAGATGTTGCCTTTGAAGTCCCATTTTGTCTTTGTTTTTAAAGTCCCATCTTATTGCAGTGATT 4440
Db 80171 TAGGCTAGATGTTGCCTTTGAAGTCCCATTTTGTCTTTGTTTTAAAGTCCCATCTTATTGCAGTGATT 80230

Qy   4441 TATATTTAACTCATGACAAATCAGGCTCTCTATTCTAAGTCCAAGACATAAAAACTTTT 4500
Db 80231 TATATTTAACTCATGACAAATCAGGCTCTCTATTCTAAGTCCAAGACATAAAAACTTTT 80290

Qy   4501 ATTGTGGAATTCAGGCATCAGTAAAATCTTTTGGGTACTCACTATGTTCCTGAAAATCA 4560
Db 80291 ATTGTGGAATTCAGGCATCAGTAAAATCTTTTGGGTACTCACTATGTTCCTGAAAATCA 80350

Qy   4561 ATCTATTTGAGTGACTTCTTCCTTTTAATTAGTCTGACCTCTTTAATGTGCACTTCGACTTG 4620
Db 80351 ATCTATTTGAGTGATCACTCTTTAGTTGGTCCCAGGTAAACAAAGAAGGCCATGGTCTTTC 80410

Qy   4621 TTTGAGTGACTTCTTCTTTCCCTTTAATTAGTCTGACCTTTAATGTCAGTTCATATTCATATTTAAT 4680
Db 80411 TTTGAGTGACCTTCTTCTTTCCCTTTAATTAGTCCCAGGTAAACAAAGAAGGCCATGGTCTTTC 80470

Qy   4681 TTCATTTCCTGGTCATCTTCTTTCCCTTTAGTTGTTCATAATTGCACTTC 4740
Db 80471 TTCATTTCCCTGGTCCATCTTCTTTCCCTTAGTTTTCATATAATTGCACTTC 80530

Qy   4741 AGTTCCTTCCACACCACCACCATCAAGGATGGCTGTCAACATTCATTTGTTCTATGGTTATATAT 4800
Db 80531 AGTTCCTTCCACACCACCATCAAGGATGGCTGTCAACATTCATTTGTTCTATGGTTATATAT 80590

Qy   4801 TCAAGGAAAAGTTGCCCAGTAGCTAATCCAATAAAATGCCCTCTTAATCATTTGTTCTATGGTTATATATAAATGCTTAGGCTAGAGAC 4860
Db 80591 TCAAGGAAAAGTTGCCCAGTAGCTAATCCAATAAAATGCCCTCTTAATCATTTGTTCTATGGTTATATAAATGCTTAGGCTAGAGAC 80650

Qy   4861 TTTTTCCTATAATTTAAATGCATCTTTCTGTAGATTATGTCCCTCCTCCACCCACTTTACATTT 4920
Db 80651 TTTTTCCTATAATTTAAATGCATCTTTCTGTAGATTATGTCCCTCCTCCACCCACTTTACATTT 80710

Qy   4921 GTCTGCTTGTCTCCCTTGCTGTTGTTGTTGTTGGTGGGTAGTGTGGGA 4980
Db 80711 GTCTGCTGTGTCTCCCTTGCTGTTGTTGGTGGGTAGTGTGGGA 80770

Qy   4981 TGTTCAAGGGCGACACGTATTGGGATTGGGTAGGGCACTATGGGCATTGTGTTGCTTTTCTTTT 5040
Db 80771 TGTTCAAGGGCGACACGTATTGGGATTGGGTAGGGCACTATGGGCATTGTGTTGCTTTTCTTTT 80830

Qy   5041 ATATATTTTTGGTATTTTTGCATCTCACTGGAACCCAAACTATTTTTTCATCTTCCCACACTTAAA 5100
Db 80831 ATATATTTTTGGTATTTTTGCAACTCACTGGAACCCAAACTATTTTTTCATCTTCCCACCTAAA 80890

Qy   5101 CTATTTGATGCCTCCTGTTTCCTTATTATATAAAGTTATAGCCTTACTGATCAGGA 5160
Db 80891 CTATTTGATGCCTCTGTTTCCTTATTATATAAAGTTATAGCCTTACTGATCAGGA 80950

Qy   5161 ACCTATCTGCTTCTAAAATGAAAGCTGTTTTGGTCAGATCTAGCAATTAAATTAATTCCCTTCTT 5220
Db 80951 ACCTATCTGCTTCTAAAATGAAAGCTGTTTTGGTCAGATCTAGCAATTAAATTAATTCCCTTCTT 81010

Qy   5221 CCACTTATAGCTTTCCTCTGTAACTCTGGTGTAGGGTATTTGGTTTAGGCTATAAGATGT 5280
Db 81011 CCACTTATAGCTTTCCTCTGTAACTCTGGTGTAGGGTATTTGGTTTAGGCTATAAGATGT 81070

Qy   5281 GAAACACCCTGAAATGATTCTGTCCATGCAGGCATTCAGTCATGATAATTGTG 5340
Db 81071 GAAACACCCTGAAATGATTCTGTCCATGCAGGCATTCAGTCATGATAATTGTG 81130

Qy   5341 ATACTGATTGTCTAGGTGTTCAGAAGACACCTAATAGGGGTATTTGGTTAGTAAAAG 5400
Db 81131 ATACTGATTGTCTAGGTGTTCAGAAGACACCTAATAGGGGTATTTGGTTACAAATCAGTTT 81190

Qy   5401 GAAGGCTGGTGATACGGCAAAGCAAACTACACATATTTTCTGCCTGCTCTCTCTCTT 5460
Db 81191 GAAGGCTGGTGATACGGCAAAGCAAACTACACATATTTTCTGCCTGCTCTCTCTCTT 81250

Qy   5461 CTACAATCTCTCTTCTTCTTTCTTTATCTTTTTGAAATATCAGTTGGAGACTTAGAAGA 5520
Db 81251 CTACAATCTCTCTTCTTTCTTTATCTTTTTGAAATATCAGTTGGAGACTTAGAAGA 81310

Qy   5521 CATAAAACCCATTTGATATAAGAATGCTGTGTGTATATTCAGTTGTGGTATTGGTT 5580
Db 81311 CATAAAACCCATTTGATATAAGAATGCTGTGTATATTCAGTTGTGGTATTGGTT 81370

Qy   5581 TCCCTCGAGCGTGCCGGTTTAGACTTTTTTTACAGGAGCAGCCAGTGTGAAGGAGAAACTGTCA 5640
Db 81371 TCCCTCGAGCTGCCGGGTTTAGACTTTTTTTACAGGAGCAGCCAGTGTGAAGGAGAAACTGTCA 81430

Qy   5701 ATTACAGAATAGCAAAATCTATTTTTCCTCCACCTTCTGTTGTTGCATAACAACAAGGCAACT 5760
Db 81491 ATTACAGAATAGCAAAATCTATTTTTCCTCCACCTTCTGTTGTTGCATAACAACAAGGCAACT 81550

Qy   5761 ACTGTCTGTCATCCAGCTATTTCCTCCACCTTCTGTTGTTTGGCATAACAACCAAGGCCTTATA 5820
Db 81551 ACTGTCTGTCATCCAGCTATTTCCTCCACCTTCTGTTGTTTGCATAACAACAAGGCAACT 81610

Qy   5821 TCCCCAAATCACTGCGTGGAGACGGATGATCATCCGCCACCTTCTTTTTGAAAATCGTGAGGA 5880
Db 81611 TCCGCAAATCACTGCGTGGAGACGGATGATCATCCGCCACCTTCTTTTTGAAAATCGTGAGGA 81670

Qy   5881 TCAGATCTTGGACCATGTTATATTAATGTATATTAATCCAAAAGAGGAAAAGCCATTGGGA 5940
Db 81671 TCAGATCTTGGACCATGTTATATTAATGTATATTAATCCAAAAGAGGAAAAGCCATTGGGA 81730

Qy   5941 GTCAGCTCCTTAAGTAAGCTCCAGAAGTTCCTTCGTTACTTTCTTCCTTCAGGAAGCAACTT 6000
Db 81731 GTCAGCTCCTAAGTAAGCTCCCAGAAGTTCCTTCGTTACTTTCTTCCTTCAGGAAGCAACTT 81790

Qy   6001 CCTTGATATTTTTTTTTTTTACAGGCCATATGAATAAAAACTATATTTTCAGCAGCATTGTACAC 6060

This page contains a genomic DNA sequence alignment (Qy/Db query–database comparison) consisting of paired nucleotide rows with match bars (|) between them, numbered by coordinate positions.

```
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
AUTHORS     Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
            Unpublished
REFERENCE   2  (bases 1 to 191699)
AUTHORS     Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
TITLE       Direct Submission
JOURNAL     Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -----------------------------------------------------------
            Center: Stanford DNA Sequencing and Technology Development
            Center
            Center code: SDSTDC
            Web site: http://sequence-www.stanford.edu/group/human/
            Contact: hum-info@sequence.stanford.edu
            -----------------------------------------------------------
            Project Information
            Center project name: 880
            Center clone name: RP11-288018
            -----------------------------------------------------------
            Summary Statistics
            Sequencing Vector: M13mp18; X02513; 100% of reads
            Sequencing Vector: plasmid; plasmid_accession; 0% of reads
            Chemistry: Dye-primer; 1% of reads
            Chemistry: Dye-terminator Big Dye; 99% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 190680 bases at least Q40
            Consensus quality: 191287 bases at least Q30
            Consensus quality: 191336 bases at least Q20
            Insert size: 195548; agarose-fp
            Insert size: 191499; sum-of-contigs
            Quality coverage: 7.9x in Q20 bases; agarose-fp
            Quality coverage: 8.1x in Q20 bases; sum-of-contigs
            NOTE: This is a 'working draft' sequence. It currently
            consists of 3 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *        1    12646: contig of 12646 bp in length
            *    12647    12746: gap of unknown length
            *    12747    9491: contig of 8221 bp in length
            *    9962    9061: gap of unknown length
            *    95062  191699: contig of 96638 bp in length
FEATURES             Location/Qualifiers
     source          1..191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288018"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1..12646
                     /note="assembly_name:Contig5"
     misc_feature    12747..9491
                     /note="assembly_name:Contig6"
     misc_feature    95062..191699
                     /note="assembly_name:Contig7"
```

```
ORIGIN                          clone_end:SP6"

            Query Match        99.6%;   Score 8009.2;  DB 2;  Length 191699;
            Best Local Similarity  99.8%;  Pred. No. 0;
            Matches 8028;  Conservative 0; Mismatches 13; Indels 1; Gaps 1;

Qy       1 CAAAATTGAAGTCCTGGGCAGCGTTACAAAACTAGAGGTCATAAAAGTTTACCCTAATTACC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8828 CAAAATTGAAGTCCAGGGCACGTTACAAAACTAGAGGTCATAAAAGTTTACCCTAATTACC 8769

Qy      61 AAGAGATTTCCTAGAGGATCTATATAATTGGAATCCAGAATTCCAGAATCTCGCCCTCTGTAAAGTTCAAGCA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8768 AAGAGATTTCCTAGAGGATCTAGAATTGGAATCCAGAATTCCAGAATCTCGCCCTCTGTAAAGTTCAAGCA 8709

Qy     121 CTTTCCATGACACCATACTGTTTCTTTCCACCTGCACAATGCAAAATGAAACTCTTATGAAA 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8708 CTTTCCATGACACCATACTGTTTCTTTCCACCTGCACAATGCAAAATGAAACTCTTATGAAA 8649

Qy     181 CTGCTGTTTCTTATCCTGGGCTAAATGTTGCAGAAAAAAGATTTAATCTTTGGGATAAGGC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8648 CTGCTGTTTCTTATCCTGGGCTAAATGTTGCAGAAAAAAGATTTAATCTTTGGGATAAGGC 8589

Qy     241 TATTTTGGGTTTCTCCCTACTTCTTGGGAAACAAAGGTTTTCTTCCCTGGGCTAATTAAGT 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8588 TATTTTGGGTTTCTCCCTACTTCTTGGGAAACAAAGGTTTTCTTCCCTGGGCTAATTAAGT 8529

Qy     301 GTGGTATTGTTCTTCCCAGGGAAATCAGTGATGCATCACCTGTGCTATCAAATGTCAGGG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8528 GTGGTATTGTTCTTCCAGGGAAATCAGTGATGCATCACCTGTGCTATCAAATGTCAGGG 8469

Qy     361 TTGGAGTTCCTGATTATTGCATGTGCCCACAGAAGCTTGGGTGCAAAGGAATTGGACACATT 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8468 TTGGAGTTCCTGATTATTGCATGTGCCCACAGAAGCTTGGGTGCAAAGAAATTGGACACATT 8409

Qy     421 TCCCAAAAGTAAGACATACTGGGAGTCCCTGTTTACCTTCCTGGTATACACGCATCCTCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8408 TCCCAAAAGTAAGACATACTGGGAGTCCCTGTTTACCTTCCTGGTATACACGCATCCTCC 8349

Qy     481 AGCCCCATAATCTTTGCTTTTTAGTCCTAAAAATCAATAAACTCTCATTGATGTCTA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8348 AGCCCCATAATCTTTGCTTTTTAGTCCTAAAAATCAATAAACTCTCATTGAGTGTCTA 8289

Qy     541 GGCCATTGTAGTAAACAATAAAGAGAGGAGGCCTTCTGACAACTGACAAAATTGTC 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8288 GGCCATTGTAGTAAACAATAAAGAGAGGAGGCCTTCTGACAACTGACAAAATTGTC 8229

Qy     601 ATCTGAAGTGGGTGCCAAGCAACAGCCTGGGGCTGAGCCTTGGCTACATCCTGCCCAAGTGG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8228 ATCTGAAGTGTGCAAGCAACAGCCTGGGGCTGAGCCTTGGCTACATCCTGCCCAAGTGG 8169

Qy     661 AGGATCAGTGCCCCATTTAAACATCTGGTAGAACTAAAGAACAACGCCTGCCACAATGA 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8168 AGGATCAGTGCCCCATTTAAACATCTGGTAGAACTAAAGAACAACGCCTGCCACAATGA 8109

Qy     721 CTTATTTCCCTGCATTTGATACCGTCAATCCTTGAGAAAGCTATTGTCAACAAT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8108 CTTATTTCCCTGCATTTGATACCGTCAATCCTTGAGAAAGCTATTGTCAACAAT 8049

Qy     781 AGCAAAAGGTTGGAAAAAATTTGAAATTTTACCTAGAGACACACACATTCACATCTGCTG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8048 AGCAAAGGTTGGAAAAAATTTGAAATTTTACCTAGAGACACACATAGTTCACATCTGCTG 7989

Qy     841 TGTGGGCTGAATGTCTGCCCCAGTAGAAAAACAGTTCTTCTTAAAGCTATTGTCAACAAT 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7988 TGTGGGCTGAATGTCTGCCCCAGTAGGAAAAACAGTTCTTCTTAAAGCTATTGTCAACAAT 7929

Qy     901 ACCTTCCAGATGTTAGCATTTTACAATTTAAGGAAACTTAAAATAGCCTTCAAAACTTTTTG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7928 ACCTTCCAGATGTTAGCATTTTACAATTTAAGGAAACTTAAAATAGCCTTCAAAACTTTTTG 7869

Qy     961 CCAGTTTCTCTGATATCCAAATCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    7560 TCTGACCTTCACGCTTAATTTCTTCGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTA 7619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1268 TCTGACCTTCACGCTTAATTTCTTCGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTA 1209

Qy    7620 ACTTGCATATGAAAATGAACCCTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTG 7679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1208 ACTTGCATATGAAAATGAACCCTACAAAAATACAAGAGTGTCAGGATAAAGTTATGCTCTG 1149

Qy    7680 GTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTGTCCTTCAAAGAGTTAAG 7739
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1148 GTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTGTCCTTCAAAGAGTTAAG 1089

Qy    7740 CAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTCTCTTTCCCTTTCCTCC 7799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1088 CAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTCGTCTTCCCTTTCCTCC 1029

Qy    7800 TCAATAAATTCCAAAACATCTTTAGTGTCCAAACATCTTTACCAGTGCCCTGAGGCTCCAGAG 7859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1028 TCAATAAATTCCAAAATCATCTTTAGTGTCCAAACATCTTTACCAGTGCCCTGAGGCTCCAGAG 969

Qy    7860 GAGCCAATTGCCTTCTTCTCAAGGTCACATAGGGTGGGGTGGTTAGGACCCAAAATCTAGAATT 7919
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    968 GAGCCAATTGCCTTCTTCTCAAGGTCACATAGGGTGGGGTGGTGAGTTAGGACCCAAAATCTAGAATT 909

Qy    7920 CCTGACTTCCAGTAACTTCTCGAAGTGCATTTTGTTTCATTTTGGTTTTTATTATTATAGA 7979
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    908 CCTGACTTCCAGTAACTTCTCGAAGTTCATTTTGTTTCATTTTGGTTTTTATTATTATAGA 849

Qy    7980 ATACTTGCTAAGCACACTAACCCCTGCATTGATTAATAACTTAGGGATTCTCAGGTGGAT 8039
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    848 ATACTTGCTAAGCACACT:ACCCCTGCATTGATTAATAACTCTAGGGATTCTCAGGTGGAT 789

Qy    8040 CC 8041
           ||
Db    788 CC 787
```

```
RESULT 4
ACI25563
LOCUS       ACI25563/c         275631 bp    DNA    linear    HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9B12, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   ACI25563
VERSION     ACI25563.4  GI:24817906
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 275631)
  AUTHORS   Mizny,D.Marie, Metzker,M.lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,M., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
```

```
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,Q., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,G., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pai,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,V., Popovic,D., Primus,E., Pu,L-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Snajs,D.,
            Sneed,A., Sodergren,E., Song,X-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 275631)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 275631)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23095552.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----------------- Project Information
            Center project name: GDBK
            Center clone name: CH230-9B12
            ----------------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 228152 bases at least Q40
            Consensus quality: 230310 bases at least Q30
            Consensus quality: 231872 bases at least Q20
            Estimated insert size: 235029; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
```

```
        *  consists of 4 contigs. The true order of the pieces
        *  is not known and their order in this sequence record is
        *  arbitrary. Gaps between the contigs are represented as
        *  runs of N, but the exact sizes of the gaps are unknown.
        *  This record will be updated with the finished sequence
        *  as soon as it is available and the accession number will
           be preserved.
        .
        *  1       192878: contig of 192878 bp in length
        *  192879  192978: gap of unknown length
        *  192979  270295: contig of 77317 bp in length
        *  270296  270395: gap of unknown length
        *  270396  272405: contig of 2010 bp in length
        *  272405  272505: gap of unknown length
        *  272506  275631: contig of 3126 bp in length

FEATURES             Location/Qualifiers
     source          1..275631
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9B12"

     misc_feature    3010..30884
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH339524"

     misc_feature    complement(55153..56817)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH339528"

     misc_feature    57750..59243
                     /note="wgs_end_extension
                     clone_end:Sp6"

ORIGIN
      Query Match        5.3%;         Score 426;  DB 2;    Length 275631;
      Best Local Similarity  49.0%;    Pred. No. 2.2e-78;
      Matches 1940;  Conservative  0;  Mismatches 1770;  Indels 246;  Gaps  47;
```

```
Qy    3672  TTAAACAAATGAACAAAATAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTTACCA  3731
            ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  145145  TGAACAGAGTGAACGTGTCACAAAAGACCCTTTAAGCAAATCCCATCCAC--ATAA  145089

Qy    3732  GAGTCATTCAGGAATGACAAATCATAGAATAACAGAAATTGATGCTTGTGCATATCAGA  3791
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  145088  GAATCAATGCAGGAATGCAGAAAACCAGCAGATAAATAACGGTGGCTATCAGA  145029

Qy    3792  GAAAAGAAGTGGAAGAGGTTGTTCAAGGTAGTATCATGATGTACCAGTCCTCCCTCAAACAC  3851
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  145028  GAGGGAGGGGCTGTTGTCAGGGTATCATGGCTCACTGGT-CTCGGCCTCCTC-AACAC  144971

Qy    3852  AAATCTGCAAGTCCCACAGTGAAAAAGTAAGTTAACTCAATGTGAAGCGGTTTTACAAACACT  3911
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  144970  AAGCAGCCAGTCGCA-TTAAAAAAAAGGCTAATTTATGAGAAGCTTTCAAACTATCA  144912

Qy    3912  TTTTT-----AAAAGTCTTAAAACTCCTAAGAAAAGCAAGAGATTAATTAGTCAAAGAAGTG  3965
            |||||                                      ||||||||||
Db  144911  GGAGTGCAAGGGACCAGGCAAAGAAGTGAGTGATTTCTTTGGGGAAGCAGTC  144852

Qy    3966  AGTAAACATGAAAATGCCTGAACACAGTAATGA--GCTAAGCACAAAAGTTAGAGACATGTT  4023
            ||||||||||||||||||||||||||||||||||    |||||||||||||||||||||
Db  144851  ATGCTGTTAGCTGTATAGCCCCTGCTTGCATGGAATAAGAGGAATCCAGCAGAGGAGAGA  144792

Qy    4024  AGTTAATATGTTCTTGAAAGCAGCACTCCTTGCATGGAATAAGAGGAATCCAGCAGAGAGGGGCAA  4083
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  144791  GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA  144732

Qy    4084  GTGAACACTCCTTGAATAAAATGTAAAAATTAATTAATTTTGGGTTTATGTTCTATACTGTGTA  4143
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  144731  GAGATCCTTGACTGAAATGAAATAAATTAATTAAATATTCTCTGCTAGTGTA  144672

Qy    4144  TAATAGAATGAAATGATTAAAAATTTATTTGTAGTTTAGAAAATAATCCTATTTAC  4203
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db  143660  TGTTGAGACACTGAAGTAAAACCTCCCTTCACCACTCCCCAGTGTGAACACGGT  143601
```

```
Db  144671  TATTCTGACAAATACAAA-TATTTGCCAGGACACTTTACAGTTTAGGAATATCTTAGTAC  144613
            |||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Qy    4204  ACAGTTTACCTTAATTGATTGAAGACTGTTTGAGTGATGGTGGTAGCACAATCCACA  4263

Db  144612  ACAGGCGTAATCCCCTTGATAAGAGTACTACAGGGCAGCAGTCTCGTAACATCTTACA  144553
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    4264  TAACTGAGATATCAGAGACACCGTATCGGACCAGCTCTGTTGAGTGATGAAAGCT-GTAACC  4322

Db  144552  CAGCCCAGTGTCAAGACGGCCTGGGTGCCCGACCCTGTCCGGTTAACAGAAAATCCGTGTC  144493
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    4323  TCAGGAAGTCACTTTCTCTTTCTCGGGCTTCTCAATTTCCTTTTTGGGATGAAATGAGAGAGTGTTA  4382

Db  144492  CTCGGGACGGGGACACGGTTGCTTCTTCAGTCTCTGTGTAATTCTATTTATCATCGGATCGGATAAGAAAG  144433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    4383  GGCTAGATTGCCTTTGAAGTCCCATTTGTCTTTAAAGTCCCATCTATTGCAGTTA  4442

Db  144432  GCAATAGA------------CCGAGTGGTTTCAAAGTCCATTTTAGCGC------A  144395
            ||||||||            ||||||||||||||||||||||||      |
Qy    4443  TATTTAACTCATTGACAAATCAGGCTTCTCTTATTCTAAGTGAAGACATAAAAACTTTTAT  4502

Db  144394  TTTAGAAACTGAATGCCACACCTCTTTTCTTCTGGTTCTCAAAATCCAAGGTTATCAGTTAGAAAG  144335
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    4503  TGTGGAATTTCAGGCATCAGTAGTAAATCTTTTTTGGGTACT-CACTTATGTTCCTGAAATCAA  4561

Db  144334  AATTGTTTTTTTTCAAGCACCAGTACAACTTTTTGAGTCTTCACAGTGTCCCTGAAATCAG  144275
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    4562  TCTATTTTGAGTGATCACTCTTTTAGGTGCCCAGTAAACAAAAGAGGCCATGGTGTCTTTCT  4621

Db  144274  ATTGTTCTGAGTGATCCCTGTCTTTAGGTGCTCGGGTAAACAAAGATTACTCTAG-------  144222
            ||||||||||||||||||||||||||||||||||||||||||||
Qy    4622  TTGAGTGACCTTCTTTTCCCTTTTAAATAGTCTGACCTCTTTAAAGTCAGTTCGACTGAT  4681

Db  144221  -AAGGTAACAGTCTCTCCCCTTTCAGTTGAT----CGCCTTGATGCAAAACTCTGCCCGA-  144168
            |||||||||||||||||||||||||||||    |||||||||||||||||||
Qy    4682  TCATTTCCCTGGTCCATCTTCCTTGGTCTGGAGGGCCTTCCTAGTTTCAATATTGCACTTCA  4741

Db  144167  --------ACCCTAGTCCAAGGTCTCCCACTTCATTGTTCTATGTTTATATAATT  144127
            ||||||||||||||||||||||||||||||||
Qy    4742  GTTCTTCTTCCACACCATCACAAGGATGGCTGCAGAACAATTCATTTGTTCTAGATTCGAAAG  4801

Db  144126  ATCCCTTCCACTCCACCACCACCGGGACAGTTGTATCAATTCCTAGGTTTCAGTGTTTGAGAAG  144067
            ||||||||||||||||||||||||||||||||||||||||||||||
Qy    4802  CAAGGAAAGAGTTGCCCAGTAGCTAAATCCAATAAAATGCCTCTTATGGGGCGGCTTAGACT  4861

Db  144066  CA--GGAGAGGTTTTCAGTGGCTGAATGTAATAGGTGGATTTAGTCTTTACAGGAGAGAGC  144009
            ||  |||||||||||||||||||||||||||||||||||||||||||||||
Qy    4862  TTTTCCTATAATTAAATGCATTCTTCGTGAGATTATGGTCCCTCCACCACTTTACATTTG  4921

Db  144008  TGGGGACTTGCCCCTTAACTGACATACC--------TATTCCCCTCTTTAACATTG  143961
            |||||||||||||||||||||||||        ||||||||||||||||
Qy    4922  TCTGTGTTGTTCCTTGCTCTGCTAGTCATGGAAGCTGTGTTG----GTAGTGGGGGCCAGTGTG  4977

Db  143960  TCTACTGTCTGCATACTACAGCTAGCCAGGGGAAGAGATGTGAANAGTAAGGTAAGGTAGGTTAG  143901
            ||||||||||||||||||||||||||||||||||||||||||
Qy    4978  GGATGGTTCAAGGGCACGTAATTGGTAGGGCCAATTGGGCTTGGTGCATTCTT  5037

Db  143900  GGCAATAAGTCTTCCTGGGTCCAGACCATAGAAACAATTGCATGATTCTATATTCTCCTTTTTATA  143841
            ||||||||||||||||||||||||||||||||||||||||||||||||
Qy    5038  TCTAIATATTTTGG------------TATTTTGCATTCCACTCGGAACCCAACTATTTT-  5082

Db  143840  TTCTTAGTACCTGAGAGAAACCCATCTCTAAAGAGAAAGGGACTGGTAACCTTCCTTCTTGT  143781
            ||||||||||||||||||||||||||||||||||||||||||
Qy    5083  ----------TCATCTCTTCACCTAAAACTAITTGATGCTTATATA-TAA  5130

Db  143780  GTGCAAATCATCTGATATCTTGTTTTTACATCGTCACCCCTGAGATTGAAGACAG  143721
            |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    5131  AGTATAGCTCACTGTAGCCTTATGATCAGGAGGAACCTATCTCGTTTCTAAATGAACTGTTT  5190

Db  143720  TGAGTGTCCACAATTCCAGGCCACTGGACTCATCTTCAGGGTGATGAGAGCTGTTT  143661
            ||||||||||||||||||||||||||||||||||||||||||||||||
Qy    5191  TGGGTCAGATCTAGCAAATTAATTCCCTTCTTCACTTAGTACCAGGCTGTAACTCTGGT  5250

Db  143660  TGTTGAGACACTGAAGTAAAACCTCCCTTCACCACTCCCCAGTGTGAACACGGT  143601
```

Qy / Db sequence alignment (nucleotide BLAST-style pairwise alignment with position coordinates).

Db   141541 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 141482

Qy   7442 TGTACTATTAAAACATTTACTTCCAAATATTCTGTCATTCAGGAGTGTATGGAAGTATCGA 7501

Db   141481 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 141422

Qy   7502 TGGCTTCTTTAAAATGAAGCAGGAGGGGTTCTGCCAGAGTATTCTATGAAAATAAGTT 7557

Db   141421 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAGGGT 141366

RESULT 5

LOCUS   AC115766

LOCUS   AC115766   224486 bp   DNA   linear   HTG 13-DEC-2003
DEFINITION Mus musculus chromosome 1 clone RP23-82I24 map 1, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AC115766
VERSION   AC115766.6 GI:39841191
KEYWORDS   HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 224486)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 224486)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,R., Kells,C., Larocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            Maclean,P., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,T.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
            Seaman,S., Severy,R., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 224486)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavskiy,L., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Collymore,A., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Diaz,J.S., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Ferreira,P., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Graham,L., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Hall,J., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Kamat,A., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Lindblad-Toh,K., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Macdonald,P., Meldrim,J., Meneus,L., Mihova,T., Murphy,T., Naylor,J.,
            Nguyen,C., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,

COMMENT   -------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
            Center project name: L23317
            Center clone name: 82_I_24
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *   1   71982: contig of 71992 bp in length
            *   71983   72082: gap of 100 bp
            *   72083 117517: contig of 45435 bp in length
            *   117518 117617: gap of 100 bp
            *   117618 122151: contig of 4534 bp in length
            *   122152 122251: gap of 100 bp
            *   122252 126124: contig of 3873 bp in length
            *   126125 126224: gap of 100 bp
            *   126225 163226: contig of 37002 bp in length
            *   163227 163326: gap of 100 bp
            *   163327 165826: contig of 2500 bp in length
            *   165827 165926: gap of 100 bp
            *   165927 169052: contig of 3126 bp in length
            *   169053 169152: gap of 100 bp
            *   169153 175011: contig of 5859 bp in length
            *   175012 175111: gap of 100 bp
            *   175112 224486: contig of 49375 bp in length

FEATURES   Location/Qualifiers
     source   1..224486
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1"
                     /clone="RP23-82I24"
                     /clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Query Match   5.0%; Score 402.6; DB 2; Length 224486;
Best Local Similarity   54.1%; Pred. No. 1.7e-73;
Matches 1738; Conservative   0; Mismatches 1174; Indels 301; Gaps 33;

Qy   3561 TGACCTCTCTACTGAGATGAAAAACTCTACTAGAGCGGAGAGATTTATCTGCTTGTATCAG 3620
          ||||| ||||| ||||||| | ||||||||||||||||||||||||| |||| ||||||
Db 128421 TGCCTCCCCTGTTCTTCCGGTTCCCTAGCAAAACTACCTCCCATCTCACTTTTTGCCATTCGTAA 128480

Qy   3621 GTACTGCTTCAAACAGCACCTGATTACAGAGTAGGTGCTCAAAAGAGATATTTCTTTAAACAA 3680
          |||||||||||||||||||||||| ||||||| |||||||| |||||||||||||||||||
Db 128481 GCTTAAGTAAATAAAATCTGTGCTTATTCACTGTGTGAGAGATGCTCAGGTGTGTCTCGAAG 128540

Qy   3681 TGAACAAATAAAAAGTAGATCTTTTGAGAGAAAAGCTCTTCCACACTTACCAGAGTCATTC 3740
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db 128541 TGAGCTTGCACAGAAGACACCTTTGGAGCAAAATCCCATTCAC---ATAAGAAGCAAGC 128597

Qy   3741 AGGAATGACAAAATCATAGAATAACAGAATTTGATGCTTTTGTGCATATCAGAGAAAGAGG 3800
          |||||||||||| ||||||||||||||||||||||||||||||||| |||||| |||||
Db 128598 AGGAATGACAAAACCATCAGATAATGGGCCATTGGTAATTGTCCCTAICAGGGGAGGGGAGG 128657

Qy   7502 TGGCTTCTTTAAAATGAAGCAGGAGGGGTTCTGCCAGAGTATTCTATGAAAATAAGTT 7557

Qy 3801 TGGAAGGTTGTCAAGTATCATGATGTACCAGTCCTCGCCTCCTCTCAAACA----CAATC 3855
Db 128658 TGAGCCATTGTCAGGGTATCCTGAGTGTACCAGTCCGGCCTCTTCAACACAGACCAGT 128717
Qy 3856 TGCAAGTCCCACAGTGAAAAGTAAGTTAAGTACTCATGTGAGCGTTTTACAAACATTTTT 3915
Db 128718 TTCATGTAAAAAATAAAAAAATAAAAGCTAATTTATATAGAGAAGCTTTTTCAAAATA---TTCGG 128774
Qy 3916 TAAAAGTCTTAAAAACTCCTAAGAAAAGCAAGAGATTTAATAGTCAAAGAAGAGTAAAC--- 3972
Db 128775 TGGAAGTATTTAGAACTGTTCGGGAATGCAAGGGACAGTCAGTAAAGTGAGTGATAGACTTG 128834
Qy 3973 ------ATGAAATGCCTGAACAGAGTAATGAGCTAAGCACACAAAGTTAGAAGACATGTTA 4024
Db 128835 GGCAGATGAAGAGTTGAAGAAACTGGTTTTCTTGGGAAGCAATCTGCTATAG 128894
Qy 4025 GTTAATATGTCTTGAAAGCAGCAGCAGTCCTGCTTTCAAGGAGCAGAGA---------- 4075
Db 128895 CTCCCGTTTGCCTGGAACAAGGAGGAAATCCAATGCAGAGAGGGGAAAGAGAGGACATGTTA 128954
Qy 4071 -------ACAAATTGGGCAAGTGAACACTCCTTGATATAAAATGTGTTAAAATTAATTTT 4121
Db 128955 CAGAGAGACAGAGATACTCACAGAGATTCTTGGCTGAAGTTATGAACAGAGAGAG 129014
Qy 4122 GGGTTATGTTCTATACTGTGTATAATAGAATGAAAAAATTATTTGACTAGCACTTTGTA 4181
Db 129015 GAGCTTTATTTATCTGCCTACTCTATGTTCTGACAATAAAAATATTTTGCCAAGCACTTCGCA 129074
Qy 4182 GTTTAGAAAATATCTCTATTTACACAGTTTACCTTATTTGATAAGACTGTTGAGTGATGGG 4241
Db 129075 GTTTAGAAAATATCTCTATGTPACACAGCATTCTTTGATATAAGAACATTGAATGATAGA 129134
Qy 4242 ATAGAGATAGGTGGACAAATCCACATAACTACTGAGGTATCGAGAAGACGCTC 4301
Db 129135 ACAGCTCTGTGGACAGTCTACACAACCCAGTGTCAAGAAACCTTGGGGTCTCGCATCCTGTCG 129194
Qy 4302 TGTTAGTAAGAAGCGTAA----CCCAGCAAGTCACTTTTCTTCTGGGTCTCTATTTC 4358
Db 129195 GGTTAACAGAAAATTCTGTATCCTTGGGGAGGCTGGCTTCAGTCTCTCTGGGATTCTATTTA 129254
Qy 4359 CTTTTGGTGAAAGAGAGTGTAGGCGCTAGATGCCTTGAAGTCCCATTTTGTCTTTTAA 4418
Db 129255 TCTTCCGGGTGAAATGAGGGGTACAGATTGAG----TGGTTTTAA 129293
Qy 4419 AGTCCCATCTCATTGCAGTGATTTATATTTAACTCATGACAAATCAGGGTTCTCTTAATTCT 4478
Db 129294 AAGTCCATTTTAGCTCAGTTTATATAGCTAAATGC-----TATACCTCTTTCCTCTAGATCCG 129347
Qy 4479 AAGTGCAAGACATAAAAACTTTTATTGTGGAAATTCAGGCCATCAGTAAAATCTTTTGGGTA 4538
Db 129348 AGATAAAAATGTTCTTCCTCGGGATTCCTTTTTCAAGCAGCGGTACATTTCTTTGAGTC 129407
Qy 4539 CTC-ACTTAGTTCCTGAAATCATCTATTTGAGTAGTGAAACTCTTTTTAGGTGCCCCAGGTA 4597
Db 129408 TTCTACTGATGTTCCTGAAACCAGATGTTCCTGAGTGCTTCCCTGTCGTGGGTGCTCAGGTA 129467
Qy 4598 AACAAAGAGAGGCCAGGTCTTTCTTGACTGACTTTTAATTAGTCTGACC 4657
Db 129468 AGCAAGAAGAAGACACTTAGGTGCACTGTGGAGAAAAGAGGCTCTCTCCTTTTCCTTTTAGTGAATCTTT 129523
Qy 4658 TCTTTAAATGTCAGTTCTGACTGATTCAGTTCCTGGTCTGGTCTGAGGGCC 4717
Db 129524 TCTTTTAATGCAAAATTCTGCCTG---------CACATCCAGTCCAGGGTC 129563
Qy 4718 TTCCTAGTTTTCATAATTGCACTTCAGTTCCTTCCACACACACCATCAAGGATAGGCTGTCAAC 4777
Db 129564 TCTCTCCACTTCATCCTGCTATCGTCATCCCTCCCCAAAGGGACCACTCTCATC 129623
Qy 4778 ATTC---ATTTGTTCTATGTTATTAATTCAAGGAAAAGTGCCCAGTAGGCTAATCCAATA 4833
Db 129624 TTTTCTCATCTAGGTTCATCTTAGAAAGCAAAGAGGAGGTTTTC--ATAGCTGAATTAATA 129681

Qy 4834 AATGCCCTCTTATGGGCGGGCTAGAGAGACTTTTTCCTATAATTAAATGCAATCTTCTGTAGA 4893
Db 129682 AGTTTTTTTGTTGTTTTGTTTTTTTTTTGTTTTGTTTTTATTTGTCTTTTACGAGGAGAG 129741
Qy 4894 TTATGGTCCCTCCACCACTTTACAT------TTGTCTGCGTCTCCTTGCTCTAGTCA 4948
Db 129742 CTGGGGGACTTTCCCTATACTGACATACCTATTGCCCTCTTTTACGTTTGTCTGTTGCTATCTC 129801
Qy 4949 TGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGGGTAGGGCC 5008
Db 129802 CATACTCAGCTAGCCTCAAGGAAGATGTGTTACGGATAAAGGAGGCAGGTGTTTGGGTCCAGAGC 129861
Qy 5009 ACATATGGGGCGATTGCCTTTGTGTGCCATT---CTTTCTATATTTTTTTTGT 5052
Db 129862 ATAGAAGAACATTACACTGCTTCCTATTTCCCATTCCTATATTTTTTGA------- 129921
Qy 5053 ------------------------TTTTGCATCTCACTGGGACCCAA 5075
Db 129922 CCCTTCTCTAAAGAAAAGCTATTGGTACCTCTCTGTCCGGAAAAATC 129981
Qy 5076 GTTTTTATTTACTTTCAGCCCCCAAAGAAGACTGAAAGAGAGTGTTTTTGGT 5134
Db 129982 CTATTTTCATTCTCTTCCACCTAAACTATTTGATGCCTCTCTGTCTTAAAATGAAAAGCTTC 130041
Qy 5135 TAGGTCACTGTAGCCTATGATCAGGAGAACCTATCTGCTTTGTCTAAAGCAAAGCTGT 5194
Db 130042 TAGGTCACTG---GATTAAGACCAAGGCACCTGTGTGACTTAAGGCGAAAGCTGTTTTTGGT 130099
Qy 5195 CAGAGTCTACCAATTAATTCCCTTTCTTCCACTTATAGCTTTCCTCTGTA------- 5245
Db 130100 CAGACACTGAAAGCAAAACCTCTTCTTCCCAGTCACAACCTGTGTGTACACAGCCAAT 130159
Qy 5246 CTGGTCTGGTATTTGTTGGTTTATGGGCTATAAGAGTGGAAAACACCTGAATGATTCGTCCAT 130159
Db 130160 TTGCGTTGCCTATCTGGGTATTTCAACCATGCAAGAGTACTTGCAGGAGT 130219
Qy 5306 GCAGGCCATTCAGTTCAGTGAATTGTAGTAGAAAGATACTGAATGTGATACGGCAAAGCAAA 5365
Db 130220 CTGAAGAGATTTCAGTCCGTGAGTGAATGTGAAAGATGTCT----TGGAAAACTCAGAA 130272
Qy 5366 ACACCTATAGGGCTAATANTTCTTACAACTCAGTTTGAAGGGCTGGTGATACGGCAAAGCAAA 5425
Db 130273 GCCACAGCGACGGGATNGGCTCTCCCACGGTTTCACGAGCCTGTTAATATGCAAGGCAAA 130332
Qy 5426 CTACATATTTTTCTTGCCTGC------TCTCTCTTTTCTCTCTTACATCTCTCTTTCTTT 5478
Db 130333 CTTACACATCCTTTCTGCCTGGAATCTTCTCTCTCTCTCTCTCTCTCTCT 130392
Qy 5479 ATCTTTTTGAAATATC------------------AGTTTTG 5499
Db 130393 CTCTCTCTCTCTCTCTCTCTCCCCCCCCCCCCTCTCTTTTTGAATAAATA 130452
Qy 5500 CTCATCTACTTCCCTCCTTTGGGTTCCTCGGGTTAGAGCTTGGCCGGGTTAACCCAGG 5559
Db 130453 AAGATTTACCATTACATATACATACAAACTGAAATTGGCATTGGTGTTTAATGTG 130512
Qy 5560 CTCATCTACTTCCCTCCTTTGGGTTCCTCGAGCTGCGGGTTAGACTTTTTACAGGACGCCAGG 5619
Db 130513 TCCATCCACCCCTGCTTTGGTCCTTTTAGTC------ 130559
Qy 5620 CATGTGAAGGAGAAAACTGTCAGTCGTTCTAGGCTGAATTCGTTGTTTCACAAGAATTTCTAGAA 5679
Db 130560 GGTCTTCAGGAGGAGAAACGCAAAGCTGGACTAGAGTTTTCCTAGGGTGTAGTTT 130619
Qy 5680 AAGTAATTCCTCAGTCAGTTGATTACAGCATAGCAAATCTATTTTTTCCTAGGGTAGTTT 5739
Db 130620 ATGTCT--CAAGGTCATATTGATTACAGACATACTGGATTTTAAACTTCCAT---TGGCTT 130673
Qy 5740 CTGTAATGCTGCCCGGGGGGCTTATTAACTTGTCGTCCAGCTATTTCTCCACCTTCTTGTT 5799
Db 130674 TGCTCATGTCTGCCGGACAACATATAGTCGTTTTGCTATCCA---------CATCCCTTCGTGT 130722
Qy 5800 TGCGATAACAACCAAGGCAACTTCCGCAAAATCACTGCGGTGGAGGACGATGATCCTGCCAGCT 5859

```
Qy         130723 CTACCAATCAGGCAGGATATTTCATTAAATCACTGCCAGAGACCACGGATCTTAGTAGTACCA 130782
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           5860 CCCTTTTGGAAAATCGTGAGGACTAGAGATCTGGACCATCATATAATATGATGCTTCTAATCC 5919

Db         130783 GCTCTGCCTCA--------                                          130819
                  |||||||||||
Qy           5920 --CACTACCCAGAATAATATTTCTAACCC                              5941

Qy           5920 AAAAGAGGAAAGCCATCATTTGGGAGTCAGTTCCTAAGTAGCTCCAGAATTCCTGCTGGTACT 5979
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         130820 AAGGAGAGAACATGTAATTAGAGACCCAGACGTAGAGTAGGATGAAGACTCCTTGCTGCTTTC 130879

Db         130880 TTTCTTCTTGGGAGAAAACAGCAACTTCCTTGATTGATTTTTTTACAGGCATATGAATAAACT 130931
                  |||||||||||||||||||||||||||||||||||||||||
Qy           5980 TTTCCTTCAGGAGAGCAACTTCCTTGATTGATTTTTTTTACAGGCATATGAATAAAAAACT 6039

Qy           6040 ATATTTTGCAGCATTGTACACTTTTTTTTCTTTTCTGGAAATTCTAAAACTCTGACATTG 6099
                  |||||||||||||||||||||||||||||||||||||||
Db         130932 ATTTTTTCAGCACGGTGGATTCTATATTTCCTTTCTACAAATGCAAAAATTGACGGCGTAG 130991

Db         130992 GTAGAAAACGTTAAGCAAAACCTCATTTTTCTATATGCCTACTTTGCGAATGGCTTCTCTCC 131051
                  ||||||||||||||||||
Qy           6100 GTGGGAGACAATTGAG---TACATTTTTTCCCAATATCCCTACTTTTTCAGAAAGGATTTTCTCT 6156

Qy           6157 GCTCGCTTCACTTAACATTGCTGATGCGGTCAGTGCTGTTTTCTTCCTTCACCCTTTCTTTCAGGGGCT 6216
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         131052 GCTGGGTCCCATTTGGCATCCCTGATGCGGTCAGTGTTTCTTCCTTTTTCACCCTTTCTTTCAGAGGGCT 131111

Db         131112 GGAGAGGCAGCGGCGCGCGACGAGAGCCTGCAGACCGTGATCGATGGCGTCCG 131171
                  ||||||||||||||||||||||||||||||||||||||||||
Qy           6217 GGAGAGGCAGAGGGAGAGAGACGAGAGCTGTACTTGTCGCAGGCGGTCGTCTGATTGGGCTGGAC 6276

Qy           6277 GGT-CGTAGCTGGGCGTATAAAAGAAGACCCCTAGCCAGGAAGACGCTCAGAGGA 6335
                  ||||||||||||||||||||||||||||||||||||||||||||||
Db         131172 GGTGCGGTGCCTGACGTATAAAAGAGACACCCCACGGGTTTGAAAGCTTTGAAAGCTCAGAGGA 131231

Db         131232 TTCTGACAGATATCTTTACCGGGACGCAGAAG--CRACTGTGCTCAAAAATAGAAAACCACCCGCG 131289
                  ||||||||||||||||||||||||||||||||
Qy           6336 TTCTTGACAAATATCTTTACCGGGAGAGGCAAAGTACGCTCTCAAAAGCCGGAAGCCCAAGCTC 6395

Qy           6396 CTCCTGCCCGATTTCTTTCCCTGCTTGTTGTTAAATAAGATGTGCAAAGG 6455
                  ||||||||||||||||||||||||||||||||||||
Db         131290 CTCCCGCGTGCTGGCCTTCTCCTCGGCTAGAGAATCCC--TCAGTTAAACAAGATGTGCAAAGG 131347

Db         131348 ACTTGCAGGTCTGCGGGCGTTCCTCCCCTGAGGGAGGTAAGAATTGGTTTCTGGTTGGTCAG 131407
                  ||||||||||||||||||||||||||||||||||||||||||
Qy           6456 GCTTGCAGGTCTGCGCGGGCCTTCTTGCTTGAGGAGGTAAATTGCTTTCAGCCATTAACCAT 6515

Qy           6516 ATTAAAACTTTGGCGTAGACTTTTTCAGTTATTTACACATTGTTGTACTTTACTTACCTACTTTT 6575
                  |||||||||||||||||||||||||||||||||||||||
Db         131408 GCTAAAAGTTTTGCATAGAGACTTTTTCAGTGAGTTTGCATGCTT---TTTGCTTAGGAGAAGTTCT 131464

Db         131465 CTGGGGATCAGAAAGCTATATGTCTCCAGGAAGCCC 131497
                  |||||||||||||||||||||||||||||||||||
Qy           6576 GTGCAATTAGAAAACAGTGTGGGTCAGGAGGAGGACAC 6608
```

```
RESULT 6
LOCUS       AC006581            172931 bp    DNA     linear   PRI 22-JUN-2002
DEFINITION  Homo sapiens 12 BAC RP11-259018 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC006581
VERSION     AC006581.16  GI:4914350
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172931)
  AUTHORS   Muzny,D., Arenson,A.D., Bouck,J., Chen,Z., Ding,Y.,
            Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
            Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
            Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
            Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
            Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
            Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
            Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,
            Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172931)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-1999) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 172931)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 172931)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 172931)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 172931)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 28, 1999 this sequence version replaced gi:4884002.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.
```

```
FEATURES             Location/Qualifiers
QUALSTAT-REPORT
source          1..172931
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="12"
                /clone="RP11-259018"
misc_feature    1..2001
                /note="overlaps bases 32746..34746 of clone AC006513"
                /function="clone overlap"
repeat_region   complement(371..658)
                /rpt_family="AluJo"
repeat_region   1461..1537
                /rpt_family="MER94"
repeat_region   complement(1585..1632)
                /rpt_family="MIR"
repeat_region   complement(1648..1952)
                /rpt_family="AluSx"
repeat_region   complement(1953..2243)
                /rpt_family="AluJo"
repeat_region   complement(2332..2887)
                /rpt_family="MLT1P"
repeat_region   2888..2968
                /rpt_family="MER5A"
repeat_region   3377..3424
                /rpt_family="MIR"
repeat_region   complement(3843..4148)
                /rpt_family="AluSg"
repeat_region   4988..5161
                /rpt_family="MIR"
repeat_region   5283..5404
                /rpt_family="L2"
repeat_region   complement(5870..6002)
                /rpt_family="AluSp"
repeat_region   7182..7249
                /rpt_family="MIR"
repeat_region   7260..7558
                /rpt_family="AluJb"
repeat_region   7573..7756
                /rpt_family="MER5A"
repeat_region   complement(7840..8077)
                /rpt_family="MIR"
repeat_region   8363..8472
                /rpt_family="AluSg"
repeat_region   complement(8523..8823)
                /rpt_family="AluSx"
repeat_region   8889..8986
                /rpt_family="MIR"
repeat_region   complement(8987..9299)
                /rpt_family="AluJb"
repeat_region   9300..9358
                /rpt_family="MIR"
repeat_region   complement(10435..10740)
                /rpt_family="AluY"
repeat_region   complement(10741..11016)
                /rpt_family="AluSq"
repeat_region   complement(11152..11457)
                /rpt_family="AluSx"
repeat_region   complement(12062..12361)
                /rpt_family="AluSp"
repeat_region   13188..13492
                /rpt_family="AluSp"
repeat_region   13841..14180
                /rpt_family="AluSq"
repeat_region   14215..14456
                /rpt_family="AluSx"
repeat_region   14497..14528
                /rpt_family="AluY"
repeat_region   14931..15288
                /rpt_family="(CAA)n"
repeat_region   15260..15320
                /rpt_family="AluY"
misc_feature    15260..15320
```

```
repeat_region   15413..15440
                /function="low coverage"
repeat_region   16323..16617
                /rpt_family="AT_rich"
repeat_region   17380..17432
                /rpt_family="AluSp"
repeat_region   17697..17735
                /rpt_family="(TG)n"
repeat_region   17796..17804
                /rpt_family="(TA)n"
repeat_region   complement(18345..18642)
                /rpt_family="MER33"
repeat_region   18656..18877
                /rpt_family="MIR"
repeat_region   complement(19222..19523)
                /rpt_family="AluY"
repeat_region   complement(19973..20434)
                /rpt_family="LIM4"
repeat_region   complement(20445..20569)
                /rpt_family="AluSg/x"
repeat_region   20855..21159
                /rpt_family="AluY"
repeat_region   complement(21237..21456)
                /rpt_family="Charlie1b"
repeat_region   21457..21770
                /rpt_family="AluSg"
repeat_region   complement(21771..21938)
                /rpt_family="Charlie1b"
repeat_region   complement(21975..22129)
                /rpt_family="Charlie1b"
repeat_region   complement(22200..22335)
                /rpt_family="AluSg/x"
repeat_region   complement(22390..22725)
                /rpt_family="L2"
repeat_region   22956..22976
                /rpt_family="AT_rich"
repeat_region   complement(22977..23271)
                /rpt_family="AluSx"
```

```
Query Match    4.1%;    Score 326.2;   DB 9;   Length 172931;
Best Local Similarity   75.6%;   Pred. No. 1.3e-57;
Matches 495;  Conservative   0;  Mismatches   128;  Indels   32;  Gaps   6;

Qy    1169 GCAGTAGGGCGGTGAAAACGTGTGATACANTTTCTTCCATCANTATAGGGTCGCAACCAAAA 1228
                |||| ||| | || |||||||||||||| | ||||  ||||||| | ||||||  | ||  |
Db   48094 GCTGTAAGGGTAAAAAAGTGTGATANTATNTTCCTCACCCATCNTAAGGGTTGTGGGCGGCA 48153

Qy    1229 CTCCTATAGTAAAAGACACAGGTNAATAAGACGAAAAACCTAACAAAATTTATTTAAATCAAAAGT 1288
                ||||||| |||||| ||| ||||| |||||||| ||||  | |||||||  || |  ||||| |||||
Db   48154 CTCCTATAATAAAAAGACAGGTTAATAAGACNAAAAGCATAACAAATAAGTACTTCATCAAAAGT 48213

Qy    1289 TTTACATGACATGGGAGNGCTTCAGAAATGNAGACAACCCAAAGACCCAGGGAAACTGTCTGT 1348
                ||| | ||||| | | | || ||||||||| || |||| ||||||||||| ||||| ||| |
Db   48214 TTTAGGTGACACCAGTGTCTTCAGAAATGAGAAGAACCAAAAGACCCAGAGAACACTGT---G 48270

Qy    1349 TTTTTTTGCTGAGGTTCGGATGAAGAANTGGATNAGCATGTNAGCGGTTGTGAGATTAGACAAAAG 1408
                | | ||| |||| ||| | |||||||| |||| | || ||| || |||| || | |||||||||||
Db   48271 TCTGTGTCGTTAGGTCTCATGAAGAANTGGGCACGCGGTGTAGGAATGTGATTGGACAAAAG 48330

Qy    1409 GATATGAGTCTAGTGGGTAAAGGACTCAG--GGGGAAACACAGCAAGGCCTGTGTCATTCAGAT 1467
                | |||||||||| |||||||| |||| || ||| |||| ||||||||  | ||| ||||||||
Db   48331 GGTATGAGTCTANTGGGTAATGGACTAAGCGGGAAACCCAGCAAGGTTCTCTCGTTCAGAT 48390

Qy    1468 TCTTCTTGANTCTCTCTCNTANGTATAGCANTTCTTTCCTCTCGAGTATGGGGCCAGGAC 1527
                 |||||||| | ||||||   |  |  |||| |||||||||||||  | |||||||||||
Db   48391 CCTTCTTGGCCCTCTC------TGTCCAGCATTCTTTCCTCTCGGCTTTGGGGCCAGGAC 48442

Qy    1528 TCTTCTTTCAAGTACGGGGTGCTTCAAGGGAGGAAAGGGAGAAAGTGGCCCTTTTTAGATTTTATGG 1587
                 ||||||  || | | |||||| ||||  ||| || |  | ||| |||     ||| || ||||||
Db   48443 CCTTCTTCGGAAGTGAAGGGGTGTTCAANTGGCAACAACAGAGAAAAAGTGACNTTTCTAGGTTTTAT 48502

Qy    1588 CTTGCCTTCGGGGAAGAGGAGTTCTAGTTTCTNTATGACCCATCTTGGGGAAGGAAGAGGAANTTCTG 1647
```

RESULT 7
AC119619/c

```
Db  48503 CTTGGTTGTTGGGGAGGAAGTTCTAGTTTCTCTGACCAACCTTGGGGAAGGAGGAATTATG 48562
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1648  GTTTCTTGTGACT-GTTCATGAAGAAGTAAGAGGAGGGGCAGGAGGGGGT 1707
Db  48563 GTTTCTGTGACT--CCTCAGGGGAGAAAAGAGAGAGGTGAGAGCAGGAGGGGCAGGAGGT 48620
          |||||||||||   || |||||||||||||||||||||||||||||||||||||||||
Qy  1708  CAGAAAGAGACTTGGCTGCGTTCTCGAGGGGTTCCCGCTCCCTTTAGTTCCAAGTACTTCTT 1767
Db  48621 CAGAGA------CTTTGCTTCTCAGG 48662
          |||||        ||||||||||||
Qy  1768  AGCATACCAAAGCACTATACTTTGGTCATATGGGTTTTCTGAGCTCTAACACTGCAA 1822
          ||||| ||||||| | ||||||||||  |||| |||||||||||| | ||| |  |
Db  48663 AGCATGCCAAAGCGCCATACTTTGGGGTATCATTTTCTGAGCTCCACACAAAGGTA 48717
```

LOCUS   AC119619   94158 bp   DNA   linear   PRI 28-SEP-2002
DEFINITION Homo sapiens X BAC RP11-1J4 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC119619
VERSION   AC119619.5 GI:23343650
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 94158)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,O.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Geo,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliver,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Liou,C., Liu,J., Liu,W.,
            Louisseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,M., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Sekhon,M., Shen,H.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,P., Telford,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watling... Wu,C., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 94158)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
```

  JOURNAL   Submitted (29-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 94158)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 94158)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

COMMENT     On Sep 28, 2002 this sequence version replaced gi:22297319.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.

FEATURES             Location/Qualifiers
     source          1..94158
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-1J4"
     misc_feature    complement(1..2004)
                     /note="overlaps bases 1..2004 of clone AC112499"
                     /function="clone overlap"
     repeat_region   4..92
                     /rpt_family="L1MB2"
     repeat_region   complement(479..781)
                     /rpt_family="MER58B"
     STS             1591..1925
                     /standard_name="SHGC-111221"
     STS             1695..1925
                     /standard_name="SHGC-68548"
     repeat_region   complement(1803..1861)
                     /rpt_family="MER91A"

repeat_region complement(2158..2465) /rpt_family="Alusx"
repeat_region 3075..3245 /rpt_family="MIR"
repeat_region complement(3246..3436) /rpt_family="LTR54"
repeat_region 3437..3739 /rpt_family="AluJo"
repeat_region complement(3740..4007) /rpt_family="LTR54"
repeat_region 4008..4046 /rpt_family="MIR"
repeat_region complement(4405..4640) /rpt_family="MER30"
repeat_region 4834..5153 /rpt_family="MER4D"
repeat_region complement(5151..5592) /rpt_family="L1M4"
repeat_region complement(7670..7719) /rpt_family="L2"
repeat_region complement(7872..8182) /rpt_family="AluSg"
repeat_region complement(8525..9084) /rpt_family="MLT1E"
repeat_region complement(9362..9650) /rpt_family="MLT1A1"
repeat_region 10392..10501 /rpt_family="(TGGA)n"
repeat_region complement(11028..11174) /rpt_family="MER5A"
repeat_region 11717..11808 /rpt_family="MIR"
repeat_region 11937..12248 /rpt_family="Charlie4a"
repeat_region 12249..12463 /rpt_family="MER20"
repeat_region 12464..12610 /rpt_family="Charlie4a"
repeat_region 12611..12947 /rpt_family="THE1B"
repeat_region 14422..14913 /rpt_family="THE1B"
repeat_region 14914..14947 /rpt_family="Charlie4a"
repeat_region 15035..15204 /rpt_family="MIR"
repeat_region 15205..15423 /rpt_family="MER58A"
repeat_region 15424..15519 /rpt_family="MIR"
repeat_region 15620..15661 /rpt_family="(TA)n"
repeat_region complement(16648..16619) /rpt_family="MER96B"
repeat_region 16639..16659 /rpt_family="AT_rich"
repeat_region complement(17007..17274) /rpt_family="AluJb"
repeat_region 17441..19201 /rpt_family="L1PA5"
repeat_region complement(19202..19496) /rpt_family="AluSp"
repeat_region 19497..23797 /rpt_family="L1PA5"
repeat_region 22116..22396 /rpt_family="L1PA5"
STS 22258..22436 /standard_name="D11S3206"
STS /standard_name="D11S2442"
repeat_region complement(24720..24929) /rpt_family="MER30"
repeat_region complement(25741..26040) /rpt_family="AluSx"

---

Query Match 3.9%; Score 317; DB 9; Length 94158;
Best Local Similarity 75.0%; Pred. No. 1.1e-55;
Matches 482; Conservative 0; Mismatches 140; Indels 21; Gaps 6;

Qy 1175 GGGGTGAAAACGTGTGATACATTTTCTTCTTCATCATCATAAGGGTCGCAACCAAAACTCCTA 1234
Db 72007 GGGGAGAAAGGTAATGAATGCATCCTTTCCTCGGCCCATCATTAAGGGCTGACATTCCTA 71948

Qy 1235 TAGTAAAAACACAGGTTAATAAGAGACAAAAACCTAACACAAATTTATTTTTAATCAAAGTTTTACA 1294
Db 71947 TAAACCAAAGATTAGGTTAATAAGAGACAAAAAC---ATAAAATGTATTTAATCAACATTTTATA 71891

Qy 1295 TGACATGGGAGTCTTCAGAAATGAAGACCCCAAAGACACCAGGGGGGAAACTGTGTCTTGTTTTTT 1354
Db 71890 TTACATGGGAAACCTTCATTAAATGAAGACCCAAAGACCCAGGGGAAATACTGTCTT-ATTTTTA 71832

Qy 1355 TGCTGAGGTTCGATGAAGAAATGGATAGCATGTAGCCATGTAGATTAGCAGAAAAAAGAGATATG 1414
Db 71831 TGATTAGGTTCGACGAAGAAATAGACGTCATGTAGCAATGTGATTAGACAAAAGGGTATG 71772

Qy 1415 ATCTAGTGGGTAAAGGACTCAGGGGCGAAAACACACCAAGGCCTCTATTCAGATTTCTTCTT 1474
Db 71771 ATCTAAATGGTATAATGACTAAAGGGAAAAAACCCCAACCAAAGGCCTGTCTCCAGATACTTCTT 71712

Qy 1475 GATCTCTCTCTTCTTATGTATAGCCATTCTTCTCCTGAGTATGGGCGGGGGAGGACTCCTCTT 1534
Db 71711 GACCTGTCTC--------TGTAGCTTTTCCTTGGAGTATGGGCCAGGAGTCCTCCA 71660

Qy 1535 CAAAGAGGTCTTCAAGGAGGAAAGGGAGGCAGAAAATGGCCTTTTTTAGATTTTATGGCTTGCTT 1594
Db 71659 AAAATGAAGGTCTTCAAGGAGGAAGAAGGCAGAGTGCACCTTTTCTAGGTTTTAATGCCTTGCAT 71600

Qy 1595 CGGGGAAGGAGGAGTTCATGAAGAAAAGAGGAGTTAAGAGGCCAGAGGGCAGAGAGAAAG 1654
Db 71599 TGTGGGAGATGGGGTTCTGGTTCCTAGCATCTTGGGAGAGAGAGAGAAAATTCTGGTTTCTA 71540

Qy 1655 TGACTTGCTTTCATGAAGAAGAAGGGCCCAACTTAGGGGTCAGAAG 1714
Db 71539 TGACTCTCTTCCGGGACAGAGAGAGAGGCAGGACACAGAGGCAGAAGTCAGAGAG 71480

Qy 1715 AGACTTGGGTTGCTTCTGAGG--GCTTCCGCCTCCCCTTTAGTTGGTTCCAAGTACTTCTTAGCATA 1773
Db 71479 A----CCTTGCTTCTGAGGTCCTCAAATCTCTTTAGTTCAAAGTA---CTCAGGCATG 71428

Qy 1774 CCAAAGCACTATACTTTTGGCATATAGGTTTTTCTGAGCTCTAACA 1816
Db 71427 CCAAGGCACCATACTTTTGAGGTATCAAAGTCCTGAGCCTCAACA 71385

---

RESULT 8

```
TITLE   Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,M., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,Z.M., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,D., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,L., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Puiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,J.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,K., Wang,Q.,
            Wang,K., Ward-Moore,S., Warren,V., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
```
```
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 175440)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 175440)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (27-NOV-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Mar 27, 2003 this sequence version replaced gi:25696694.
```
```
COMMENT   ------------------------ Genome Center -----------------------
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          ---------------------- Project Information --------------------
          Center project name: HERR
          Center clone name: CTD-2145022
          -------------------- Summary Statistics ----------------------
          Sequencing vector: Plasmid;
          Chemistry: Dye-terminator Big Dye: 100% of reads
          Assembly program: Phrap; version 0.990329
          Assembly Center Code: BCM
          Drafting Center Code: BCM
          Consensus quality: 172016 bases at least Q40
          Consensus quality: 174734 bases at least Q30
          Consensus quality: 175344 bases at least Q20
          Estimated insert size: 168157; sum-of-contigs estimation
          Quality coverage: 14x in Q20 bases; sum-of-contigs estimation
          ----
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 19 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
```
```
FEATURES   Location/Qualifiers
     source   1..175440
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="CTD-2145022"
```
```
ORIGIN
        1   1362: contig of 1362 bp in length
        1363 1462: gap of unknown length
        1463 3030: contig of 1568 bp in length
        3031 3130: gap of unknown length
        3131 4717: contig of 1587 bp in length
        4718 4817: gap of unknown length
        4818 6158: contig of 1341 bp in length
        6159 6258: gap of unknown length
        6259 7733: contig of 1475 bp in length
        7734 7833: gap of unknown length
        7834 9184: contig of 1351 bp in length
        9185 9284: gap of unknown length
        9285 11256: contig of 1972 bp in length
        11257 11356: gap of unknown length
        11357 12606: contig of 1250 bp in length
        12607 12706: gap of unknown length
        12707 14194: contig of 1488 bp in length
        14195 14294: gap of unknown length
        14295 15764: contig of 1470 bp in length
        15765 15864: gap of unknown length
        15865 17682: contig of 1818 bp in length
        17683 17782: gap of unknown length
        17783 18813: contig of 1031 bp in length
        18814 18913: gap of unknown length
        18914 20063: contig of 1150 bp in length
        20064 20163: gap of unknown length
        20164 21187: contig of 1024 bp in length
        21188 21287: gap of unknown length
        21288 22341: contig of 1054 bp in length
        22342 22441: gap of unknown length
        22442 23780: contig of 1339 bp in length
        23781 23880: gap of unknown length
        23881 25373: contig of 1493 bp in length
        25374 25473: gap of unknown length
        25474 27379: contig of 1906 bp in length
        27380 27479: gap of unknown length
        27480 175440: contig of 147961 bp in length.
```

```
Qy    1175  GGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATCATTAAGGGTGGTCGCAACCAAAACTCCTA  1234
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  114048  GGGGCAGAAAAGGTAATGAATACCTTTTCCTCGCCCATCATTAAGGGTCATGGCTCATGGCGCATTCCTA  113989

Qy    1235  TAGTAAAAGACAGGTTAATAAGACCAAAAACCTACACAAAATTTATTTTAATCAAAGTTTTACA  1294
            ||||| |||  ||||||  |||| |||||||||||    |||||||||||||
Db  113988  TAACCAAAGATAGGTTAATAAGACAGAGAAAAAC---ATAAAAGTATTTAAATCAACACATTTTATA  113932

Qy    1295  TGACATGGGGAGTCTTCAGAAAATGAAGACCCAAAGACCCCAGGGGGAAACTGCTCTGTTTTTTT  1354
            ||||||||||| |||||||||||||||||||| ||||| ||||||||||||||  |||||
Db  113931  TTACATGGGGAACCTTCATTTAATGAAGACCCAAAGACAGCCAGGGGAATACTGTCT-ATTTTTA  113873

Qy    1355  TGCTGAGGTTCGATGAAGAAATGGATAGCCATGTAGGCTAGACAAAAAGGATATG  1414
            |||||||| |||||||| ||||  ||||| |||| ||||  ||||||||||||
Db  113872  TGATTAGGTTCGACGAAGGAGTAGTCATGCATGCATGTAGGATAGACAAAAAGGGTATG  113813

Qy    1415  ATCTAGTGGGTAAAAGGACTCAGGGGAAAACACAACAAGGCCTGTCTATTAGATTTCTCTT  1474
            ||||| |||  |||||| ||||||||||||||||| |||||| ||| | |||||||||||
Db  113812  ATCTAATGGTAATAGACTAAGGGAAAAAACCCACACCAAGGCCTTGCTTTAGATTACTTCTCTT  113753

Qy    1475  GATCTCTCTCTCTCTCTCATGTATAGCACATTCTTTCCCCTGAGTATGGGGCAGGAGGGTATG  1534
            |||||||| |  ||||  |  ||||||||||||||||  || |||| |||| ||||||
Db  113752  GACCTGTCTC-------TGTAGCTTCCTTCCTTTGAGTATGGGGGAGGAGGACTCCTCCA  113701
```

```
Qy    1535 CAATGAGGGTCTCAAGGGACGAAGGCAGAAGTGGCCTTTTTAGATTTTATGGCTTGCTT 1594
           ||||| |||||| ||||||| ||| ||  ||| ||| |||| |||||||||||||||| |
Db  113700 AAATGAGGGTCTTCAAGGGAGGAAAGGGAGACAGTGACCTTTCTAGGTTTTATGGCTTGCAT 113641

Qy    1595 CGGGGAAGAGAGTTCTTAGTTTCTTATGACCCATCTTGGGAAGAGAAGGAATTCTCGGTTTCTG 1654
           | |||| ||| ||||| |||| |||| ||||| ||||||| ||||| |||||| ||| ||| ||
Db  113640 TGTGGGAGGATGGGTTCTTGGTTTCCTACGACCCGTCTTGGAGAAGAGGAAATCTCGTTTTCTA 113581

Qy    1655 TGACTTGCTTTCATGAAGAAGAGTAAGAGGAGGGCGGAGGGGCAGGGAGAGATGGGTCAGAAAG 1714
           ||||| | ||| | || ||| ||||||||||||| | || | ||||||  ||  ||| |||| |
Db  113580 TGACTCTCTTCGGGACGAAAGAGGAGGGCAGGGACACAGAGAGGCAGGAGTAAAGTCAGAGAG 113521

Qy    1715 AGACTTGGCTGCTTCTGAGG-GCTTCCGCCTCCTCCTTTAGTTCCAAGTACTTCTTAGCATA 1773
           |      |||| ||||||||  |||||   |   | |||||||| |||||  || |||||| |
Db  113520 A------CCTTGCTTCTGAGGTCCTTCCAATCTTCTTTAGTTCAAAGTA--CTCAGCATG 113469

Qy    1774 CCAAAGCACTATACTTTGGCATATATGGTTTTTCTGAGCTCTAACA 1816
           |||| ||||  | |||||  ||||    ||  ||||||| | |||||
Db  113468 CCAAGGCACCAATACTTTGAGGTATCAAAGTCCTGAGCCTCAAACA 113426
```

```
RESULT 9
LOCUS       AC073227             154125 bp    DNA     linear   HTG 12-AUG-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-652L13, WORKING DRAFT
            SEQUENCE, 45 unordered pieces.
ACCESSION   AC073227
VERSION     AC073227.1  GI:8440039
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154125)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 154125)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (10-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA

COMMENT     -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ----------- Project Information -----------
            Center project name: H_NH0652L13
            ------------- Summary Statistics -----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 111306 bases at least Q40
            Consensus quality: 137026 bases at least Q30
            Consensus quality: 139983 bases at least Q20
            Insert size: 124000; agarose-fp
            Insert size: 149725; sum-of-contigs
            Quality coverage: 3.61 in Q20 bases; agarose-fp
            Quality coverage: 3.15 in Q20 bases; sum-of-contigs
            --------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 45 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   1169: contig of 1169 bp in length
            *    1170   1269: gap of unknown length
            *    1270   1500: contig of 1500 bp in length
            *    2770   2869: gap of unknown length
            *    2870   4193: contig of 1324 bp in length
            *    4194   4293: gap of unknown length
            *    4294   5710: contig of 1417 bp in length
            *    5711   5810: gap of unknown length
            *    5811   6984: contig of 1174 bp in length
            *    6985   7084: gap of unknown length
            *    7085   8592: contig of 1508 bp in length
            *    8593   8692: gap of unknown length
            *    8693   9797: contig of 1105 bp in length
            *    9798   9897: gap of unknown length
            *    9898  11428: contig of 1531 bp in length
            *   11429  11528: gap of unknown length
            *   11529  12866: contig of 1338 bp in length
            *   12867  12966: gap of unknown length
            *   12967  14965: contig of 1999 bp in length
            *   14966  15065: gap of unknown length
            *   15066  16845: contig of 1780 bp in length
            *   16846  16945: gap of unknown length
            *   16946  19643: contig of 2698 bp in length
            *   19644  19743: gap of unknown length
            *   19744  21949: contig of 2206 bp in length
            *   21950  22049: gap of unknown length
            *   22050  24411: contig of 2362 bp in length
            *   24412  24511: gap of unknown length
            *   24512  27410: contig of 2899 bp in length
            *   27411  27510: gap of unknown length
            *   27511  30054: contig of 2544 bp in length
            *   30055  30154: gap of unknown length
            *   30155  32447: contig of 2293 bp in length
            *   32448  32547: gap of unknown length
            *   32548  34490: contig of 1943 bp in length
            *   34491  34590: gap of unknown length
            *   34591  36874: contig of 2284 bp in length
            *   36875  36974: gap of unknown length
            *   36975  39829: contig of 2855 bp in length
            *   39830  39929: gap of unknown length
            *   39930  42666: contig of 2737 bp in length
            *   42667  42766: gap of unknown length
            *   42767  46079: contig of 3313 bp in length
            *   46080  46179: gap of unknown length
            *   46180  49377: contig of 3198 bp in length
            *   49378  49477: gap of unknown length
            *   49478  52066: contig of 2589 bp in length
            *   52067  52166: gap of unknown length
            *   52167  55853: contig of 3687 bp in length
            *   55854  55953: gap of unknown length
            *   55954  58804: contig of 2851 bp in length
            *   58805  58904: gap of unknown length
            *   58905  62776: contig of 3872 bp in length
            *   62777  62876: gap of unknown length
            *   62877  66204: contig of 3328 bp in length
            *   66205  66304: gap of unknown length
            *   66305  69872: contig of 3568 bp in length
            *   69873  69972: gap of unknown length
            *   69973  73632: contig of 3660 bp in length
            *   73633  73732: gap of unknown length
            *   73733  76814: contig of 3082 bp in length
            *   76815  76914: gap of unknown length
            *   76915  80767: contig of 3853 bp in length
            *   80768  80867: gap of unknown length
            *   80868  85695: contig of 4828 bp in length
            *   85696  85795: gap of unknown length
            *   85796  90246: contig of 4451 bp in length
            *   90247  90346: gap of unknown length
            *   90347  94050: contig of 3704 bp in length
            *   94051  94150: gap of unknown length
            *   94151  99671: contig of 5521 bp in length
            *   99672  99771: gap of unknown length
            *   99772 105092: contig of 5321 bp in length
            *  105093 105192: gap of unknown length
```

```
FEATURES             Location/Qualifiers
     source          1..154125
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-652L13"
     misc_feature    1..1169
                     /note="assembly_name:Contig11"
     misc_feature    1270..2769
                     /note="assembly_name:Contig15"
     misc_feature    2870..4193
                     /note="assembly_name:Contig16"
     misc_feature    4294..5710
                     /note="assembly_name:Contig18"
     misc_feature    5811..6984
                     /note="assembly_name:Contig19"
     misc_feature    7085..8592
                     /note="assembly_name:Contig20"
     misc_feature    8693..9797
                     /note="assembly_name:Contig21"
     misc_feature    9898..11428
                     /note="assembly_name:Contig22"
     misc_feature    11529..12866
                     /note="assembly_name:Contig23"
     misc_feature    12967..14965
                     /note="assembly_name:Contig24"
     misc_feature    15066..16845
                     /note="assembly_name:Contig25"
     misc_feature    16946..19643
                     /note="assembly_name:Contig26"
     misc_feature    19744..21949
                     /note="assembly_name:Contig27"
     misc_feature    22050..24411
                     /note="assembly_name:Contig28"
     misc_feature    24512..27410
                     /note="assembly_name:Contig29"
     misc_feature    27511..30054
                     /note="assembly_name:Contig30"
     misc_feature    30155..32447
                     /note="assembly_name:Contig31"
     misc_feature    32548..34490
                     /note="assembly_name:Contig32"
     misc_feature    34591..36874
                     /note="assembly_name:Contig33"
     misc_feature    36975..39829
                     /note="assembly_name:Contig34"
     misc_feature    39930..42666
                     /note="assembly_name:Contig35"
     misc_feature    42767..46079
                     /note="assembly_name:Contig36"
     misc_feature    46180..49377
                     /note="assembly_name:Contig37"
     misc_feature    49478..52066
                     /note="assembly_name:Contig38"
     misc_feature    52167..55853
                     /note="assembly_name:Contig39"
     misc_feature    55954..58804
```

```
      *     105193   109346: contig of 4154 bp in length
      *     109947   109446: gap of unknown length
            109447   115145: contig of 5699 bp in length
            115146   115245: gap of unknown length
            115246   121014: contig of 5769 bp in length
            121015   121114: gap of unknown length
            121115   126586: contig of 5772 bp in length
      *     126587   126686: gap of unknown length
            126687   132609: contig of 5623 bp in length
            132610   132709: gap of unknown length
            132710   138801: contig of 6172 bp in length
      *     138802   138901: gap of unknown length
            138982   146419: contig of 7438 bp in length
      *     146420   146519: gap of unknown length
            146520   154125: contig of 7606 bp in length.
```

```
     misc_feature    58905..62776
                     /note="assembly_name:Contig40"
     misc_feature    62877..66204
                     /note="assembly_name:Contig41"
```

```
Query Match  3.9%;    Score 311.2;  DB 2;   Length 154125;
Best Local Similarity 70.5%;  Pred. No. 1.7e-54;
Matches   509;  Conservative   0;  Mismatches  188;  Indels  25;  Gaps   6;
```

```
Qy      1177 GGTGAAAACGTGTGATACATTTTCTTCCATCATAAGGTCGGCAACCAAAACTCCTATA 1236
Db     73372 GGGGTAGAAAGTGTGGTACGTTTTCTTGCCCATCATAAGGGTCACAGCTGGCACTCCTATC 73313

Qy      1237 GTAAAAAGACACAGTTAATAGAGCAAAAACCTAACAAAATTTATTTAATCAAAAGTTTTACATG 1296
Db     73312 ATAAAAAGACAGGTTAACACAGAGAGAAACAATAACAAAATTTATTTAATCAAAAGTTTGTCG 73253

Qy      1297 ACATGGGACGTCTTCAGAAAATGAAGAACCCCAAAAGACCGGGGAAACTGTCTGTTTTTTTTTG 1356
Db     73252 ATACAGGAGCCCTTCAGAAATGAAGAATCCAAAAGACCTAGAAGAAAACCGGTCTG--CTTTTATG 73194

Qy      1357 CTGAGGTTCGATGAAGAAATGGAATAGCATGTAGCCATGGGTAGCAAATGTGATT-GACAAAAGG 1416
Db     73193 CTTAGTTTCAACGAAGAAATGGGACAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGA 73135

Qy      1417 CTAGTGGTAAAGGACTCAGGGGGAAACAACCATGTAGCAACCTAGCAAATGTGATT-GACAAAGG 1476
Db     73134 CTAATAATAATTAAAGACT----GAGAAACCCAGCAAGGCTGTATGTCCAGAATTCTTCTCG 73079

Qy      1477 TCTCTCTCTCTCTATTGTTATAGCATTCTTTCCTCTTGAGTATAGTGGGGCAGGACTCTTTTCTTCA 1536
Db     73078 TCTCTC------TGCATAGCATTCCTCCCTCCTAGGTTATAGGGCAGAACCCCTGTGGA 73027

Qy      1537 ATGAGGGGTTCTTCAAGGGAGGAGGAAGGGGAGAAAGTGGCGTTTTTAGATTTATGGCTTGCTTCG 1596
Db     73026 ATGAGGGGTCTTCAAGGGAGGAGAAAGAATGACTTGCTAGGTTTTAGTTTATGGCTTGCTTTA 72967

Qy      1597 GGGGAAGGAGGAGTTCTTAGTTTTTCTATGACCCCATCTTGGGGGGAAGAGGAAATTCTGGTTTCTGTG 1656
Db     72966 GGGGAGGAGGGGGTTCTAGTATCCTGTGAAAGGGGAGAGGAAATTCTGGTTTTCTAG 72907

Qy      1657 ACTTGCTTTCATGAAGAAGAGGAGTAAGCAGGAGGGGCAGGAGAGGATGGTCAGAGAGAG 1716
Db     72906 ACTCACTTTAGGGAGAGATGGAGGAGGACGGGGAAAAATGGAGGGCCAGGAAGGTCAGAGAGAG 72847

Qy      1717 ACTTGGCTGCT--------TCTGAGGGCTTCGCTCTCCCTCCTTTAGTTCCAAGTAGTACTTCTTA 1768
Db     72846 ACTTTGTTTCTGAGGCTACTTCCTGGAACTCCTTGTTCAAAGTA---CTCA 72790

Qy      1769 GCATACCAAAGCACTATACTTTTGGCATATGGTTTTCTCGAGCTCTPAACACTGCAAATCATGC 1828
Db     72789 GCAATGCCAAAGTGTCATAACTTTGCAGTATCATTTTCTGAGCCCCAGAAAGAGAAAAGCTA 72730

Qy      1829 TAAACTCCTCCTATGACCTTCGAAACATTCCACTTGCTTTTATTCTTTATGTTTGTGATGGC 1888
Db     72729 AAAATGGCTCATAATCCCACCAACACATTAAGGTTTTATACATTCAAAGTTTGAGGAGAT 72670

Qy      1889 AT 1890
Db     72669 AT 72668
```

```
RESULT 10
AC018694
LOCUS       AC018694       196832 bp    DNA       linear    PRI 09-MAY-2001
DEFINITION  Homo sapiens BAC clone RP11-563P16 from 11, complete sequence.
ACCESSION   AC018694
VERSION     AC018694.4  GI:9665213
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE   1 (bases 1 to 196832)
AUTHORS   Sulston,J.E. and Waterston,R.
TITLE   Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
PUBMED   9847074
REFERENCE   2 (bases 1 to 196832)
AUTHORS   Ali,J., Maupin,R. and Lehnert,L.
TITLE   The sequence of Homo sapiens BAC clone RP11-563P16
JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 196832)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (16-DEC-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE   4 (bases 1 to 196832)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (03-AUG-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE   5 (bases 1 to 196832)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (10-SEP-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE   6 (bases 1 to 196832)
AUTHORS   Waterston,R.
TITLE   Direct Submission
JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7 (bases 1 to 196832)
AUTHORS   Waterston,R.
TITLE   Direct Submission
JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   On Aug 3, 2000 this sequence version replaced gi:7631073.
          --------------- Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: sapiens@watson.wustl.edu
          --------------- Summary Statistics
          Center project name: H_NH0563P16
          ---------------.

NOTICE: This sequence may not represent the entire insert of this
          clone. It may be shorter because we only sequence overlapping
          clone sections once, or longer because we provide a small overlap
          between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from

FEATURES   Location/Qualifiers
     source   1..196832
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-563P16"
                     /clone_lib="RPCI-11"
     repeat_region   25..466
                     /rpt_family="L1"
     repeat_region   917..1135
                     /rpt_family="L2"
     repeat_region   1137..1361
                     /rpt_family="MIR"
     repeat_region   1386..2075
                     /rpt_family="L2"
     repeat_region   2132..4576
                     /rpt_family="L1"
     repeat_region   4616..4774
                     /rpt_family="L2"
     repeat_region   5092..5202
                     /rpt_family="MIR"
     repeat_region   5241..5572
                     /rpt_family="L1"
     repeat_region   5884..6021
                     /rpt_family="ERV"
     repeat_region   6597..6704
                     /rpt_family="L2"
     repeat_region   6078..6151
                     /rpt_family="L1"
     repeat_region   6301..6477
                     /rpt_family="MER1_type"
     repeat_region   8852..8954
                     /rpt_family="L1"
     repeat_region   8954..9034
                     /rpt_family="MIR"
     repeat_region   9411..9560
                     /rpt_family="L2"
     repeat_region   10059..10313
                     /rpt_family="Alu"
     repeat_region   10395..10599
                     /rpt_family="L2"
     repeat_region   10743..11056
                     /rpt_family="MIR"
     repeat_region   11522..11673
                     /rpt_family="Alu"
     repeat_region   13784..13928
                     /rpt_family="L1"
     repeat_region   14035..14188
                     /rpt_family="L2"
     repeat_region   14229..14459
                     /rpt_family="L1"
     repeat_region   14660..14751
                     /rpt_family="L2"
     repeat_region   14802..15365
                     /rpt_family="MER2_type"
     repeat_region   16007..16689
                     /rpt_family="MER1_type"
     repeat_region   17025..17123
                     /rpt_family="ERV1"

The region from base position 99028 to 99031 is covered by a single
m13 subclone.

The region from base position 196832 of RP11-563P16.
position 196832 of RP11-563P16.
is at base position 1 of RP11-563P16; actual end is at base
sequenced to the right is RP11-652113. Actual start of this clone
The clone sequenced to the left is RP11-484112. The clone
NEIGHBORING SEQUENCE INFORMATION:
VECTOR: pBACe3.6
(http://bacpac.med.buffalo.edu)
and coworkers at the Roswell Park Cancer Institute
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong

```
misc_feature   17033..17299
               /note="similar to EST BE005163 (NID:g8265396)"
misc_feature   17050..17299
               /note="similar to EST AA528257 (NID:g2270326) nh26a02.s1"
misc_feature   17053..17299
               /note="similar to EST AI350387 (NID:g4087593) qo39d10.x1"
misc_feature   17054..17299
               /note="similar to EST N57929 (NID:g1201819) yv61e06.s1"
repeat_region  17648..17735
               /rpt_family="MIR"
repeat_region  19011..19227
               /rpt_family="MIR"
repeat_region  21465..21556
               /rpt_family="MIR"
repeat_region  21718..21927
               /rpt_family="Alu"
repeat_region  21961..22078
               /rpt_family="Alu"
repeat_region  22071..22284
               /rpt_family="MER1_type"
repeat_region  22285..22589
               /rpt_family="MER1_type"
repeat_region  22590..22729
               /rpt_family="Alu"
repeat_region  22745..22943
               /rpt_family="MER1_type"
repeat_region  23386..23989
               /rpt_family="MIR"
repeat_region  24058..24464
               /rpt_family="MIR"
repeat_region  24466..24582
               /rpt_family="L1"
repeat_region  24583..24880
               /rpt_family="L1"
repeat_region  24881..25141
               /rpt_family="L1"
repeat_region  25149..25336
               /rpt_family="L1"
repeat_region  25948..26580
               /rpt_family="MaLR"
repeat_region  27727..27361
               /rpt_family="ERV1"
repeat_region  28278..28567
               /rpt_family="MIR"
repeat_region  28861..29094
               /rpt_family="Alu"

Query Match    3.9%; Score 311.2; DB 9; Pred. No. 1.7e-54; DB 9; Length 196832;
Best Local Similarity 70.5%; Conservative 0; Mismatches 188; Indels 25; Gaps 6;
Matches 509;

Qy  1177 GGTGAAAACGTGATAATACATTTCTTCTTCCATCATCATAAGGTCGGCAACCAAAACTCCTATA 1236
         ||| ||||| ||| | | ||  || ||| |||||||||||||  || || ||| ||| |||||
Db 25953 GGGTAGAAAAGTGTGGTACGTTTCTTCTGCCCATCATAAGGTCACACGTGGCACTCCTATC 26012

Qy  1237 GTAAAAGAACAGGTTAATAAGAGAGCAAAAACCTTAACAAAATTTAATTCAAAAGTTTACATG 1296
         ||||| |||||||||| ||||||| ||||| | | ||||||||||| ||| | |||||  | | |
Db 26013 ATAAAAAGACAGGTTAACAAGAGAGAAAACATAACAAAATTTATTTAATCAAAAAGTTTGTGCG 26072

Qy  1297 ACATGGGAGTCTTCAGAAATGAAGACCCAGGGAAACTGTCTGTTTTTTTTG 1356
         | | || ||| ||||||||||||||| || |||||| ||||||  ||||| |
Db 26073 ATACAGGAGCCTTCAGAAATGAAGAGTCCAAAGAAAACCGTCTG-CTTTTATG 26131

Qy  1357 CTGAGGTTCGATGAAGAATGGATAGCCATGTGAGCAGTTAGATTAGACAAAAAGGATAATGAT 1416
         || || || | ||| ||||||| | || ||  |||  | |||| |||||||| ||  |||| |
Db 26132 CTTAGTTTCAACGAGAAATGGACAACCATTAGCAATGTGATT-GACAAAAGGGTCTGAT 26190

Qy  1417 CTAGTGGTAAAGGACTCAGGGGGGAAACACAGCAAGGCCTGTCTCTATTCAGATTCTTCTTGA 1476
         || | |  |||  |  | | | ||||||  |||||||| || | | | |||||||||||| |
Db 26191 CTAATGATAATAGACT---GAGAAACCAGCAAGGCTTGTAGTGTTGTCAGATTCTTCTCGG 26246

Qy  1477 TCTCTCTCTCTCTATGTATATGCATTCTTTCCTCCTCGAGTGTAGGGCAGGCAGGACTCTTCTTTA 1536
         |||||       |||  |||||||||| | ||||| | | | |||||||| | | ||  | | |
Db 26247 TCTCTC------TGCATAGCATTCTCCCTCCTAGGTTATAGGGCAGAACCCCTGTGGA 26298

Qy  1537 ATGAGGGTCTTCAAGGGAGGAAGAGGGAGGAAGAAGTGGCCTTTTAGATTTTATGGGCTTGCTTCG 1596
         ||||||||||||||||  |||||||||||||||||||||||  | ||  ||||||| ||||||| |
Db 26299 ATGAGGGTCTTCAAGGAGGAAGAGGGAGGAAGAAGTGGCCTAGGTTTATGGGTCTTGCTTTA 26358

Qy  1597 GGGGAGGAGGAGAGTTCTAGTTTCTATGACCATCTGGGGGAGAAGAAGGAATTCTGGTTTCGTG 1656
         |||||||||| | ||||| ||||||  | | ||||| | |  ||||| |||||||| ||||| |
Db 26359 GGGGAGGAGGGGGTTCTTAGTTTCTGAGTAATCTGACCCAACTTGGAGAGGAGGAATTCTGTTTCATG 26418

Qy  1657 ACTTGCTTTCATGAGAAAGAGGAGTAAGAGGCAGGAGGGCCAGGAGATGGTCAGAAGAG 1716
         ||| | || ||| ||||||| || || |  |||| ||| ||| |||| ||||| | | |
Db 26419 ACTCACTTAGGGAGGACATGGGAGGAAAATGGAGGGCCAGAAGAGGTCAGGAGGAG 26478

Qy  1717 ACTTCGGCTGCT--------TCTGAGGGCTGTTTCCGGTCTTCTTCTTTAGTTCCAAGTACTTTTA 1768
         |||| | || ||        ||||||| | |||||  |||| ||  ||   |||| ||||   | ||
Db 26479 ACTTTGTTTCTTGGAGGGCTACTTCTGGACCTTCCAATCTCCTTTTGTTCAAAAGTA---CTCA 26535

Qy  1769 GCATACCAAAGCACTATACTTTGGGCATATAGGTTTTCTCTGAGCTCTAACTGTCAATCATGC 1828
         |||| | |||| | ||||||||| | |||| || ||||||||||| | | || | ||  | ||
Db 26536 GCATCGCCAAAGTGTCATACTTTGCGAGTACATTTTCTGAGCCCCAGAGAAAGAGGTCA 26595

Qy  1829 TAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTTGATGGTTGAGGAGAT 1888
         ||||   | ||| || ||  |||| ||||| |||||  ||  |  |||| || | | ||  ||||||
Db 26596 AAATGGCTCATAATCCACCACCAGACAATTAAAAGGTTTATACATTCAAGTTCAGGAGGAT 26655

Qy  1889 AT 1890
         ||
Db 26656 AT 26657
```

RESULT 11

AC009867

```
LOCUS       AC009867              176773 bp    DNA     linear   HTG 22-NOV-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-76K11 map 11, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC009867
VERSION     AC009867.2  GI:11276242
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176773)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176773)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Landers,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hafez,N., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,H., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McLaughlin,J.,
            Meldrim,J., Nicol,M., Morris,W., Morrow,J., Mychalecky,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,J.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 22, 2000 this sequence version replaced gi:5822691.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
```

Contact: sequence_submissions@genome.wi.mit.edu
------- Project Information
Center project name: L2148
Center clone name: 76_K_11
------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus Quality: 16992 bases at least Q40
Consensus Quality: 173382 bases at least Q40
Consensus Quality: 174685 bases at least Q30
Consensus Quality: 175673; sum-of-contigs
Insert size: 182000; agarose-fp
Insert size: 175673; sum-of-contigs
Quality coverage: 5.4 in Q20 bases; agarose-fp
Quality coverage: 5.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*    1     18082: contig of 18082 bp in length
*  18083     18182: gap of 100 bp
*  18183     19240: contig of 1058 bp in length
*  19241     19340: gap of 100 bp
*  19341     69145: contig of 49805 bp in length
*  69146     69245: gap of 100 bp
*  69246     71797: contig of 2552 bp in length
*  71798     71897: gap of 100 bp
*  71898     76451: contig of 4554 bp in length
*  76452     76551: gap of 100 bp
*  76552     84099: contig of 7548 bp in length
*  84100     84199: gap of 100 bp
*  84200     94128: contig of 9929 bp in length
*  94129     94228: gap of 100 bp
*  94229    102601: contig of 8373 bp in length
* 102602    102701: gap of 100 bp
* 102702    115119: contig of 12418 bp in length
* 115120    115219: gap of 100 bp
* 115220    136142: contig of 20923 bp in length
* 136143    136242: gap of 100 bp
* 136243    175508: contig of 39266 bp in length
* 175509    175608: gap of 100 bp
* 175609    176773: contig of 1165 bp in length.
```

FEATURES   Location/Qualifiers
source   1..176773
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="11"
                /map="11"
                /clone="RP11-76K11"
                /clone_lib="RPCI-11 Human Male BAC"
misc_feature   1..18082
                /note="assembly_fragment
                clone_end:SP6
                vector_side:left"
misc_feature   18183..19240
                /note="assembly_fragment"
misc_feature   19341..69145
                /note="assembly_fragment"
misc_feature   69246..71797
                /note="assembly_fragment"
misc_feature   71898..76451
                /note="assembly_fragment"
misc_feature   76552..84099
                /note="assembly_fragment"
misc_feature   84200..94128
                /note="assembly_fragment"
misc_feature   94229..102601
                /note="assembly_fragment"
misc_feature   102702..115119
                /note="assembly_fragment"
misc_feature   115220..136142
                /note="assembly_fragment"
misc_feature   136243..175508
                /note="assembly_fragment"
misc_feature   175609..176773
                /note="assembly_fragment
                clone_end:T7
                vector_side:right"

ORIGIN

RESULT 12

```
Query Match     3.8%;  Score 307.2;  DB 2;  Length 176773;
Best Local Similarity  71.5%;  Pred. No. 1.1e-53;
Matches  495;  Conservative  0;  Mismatches 173;  Indels  24;  Gaps  6;

Qy      1153 CTTCTCTTCTCAAAAGTCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATA 1212
             |||| |||| | ||||| || ||| || ||||| ||||||||| | ||||||| |  ||||||
Db    104797 CTTCTGTACCTTGGAGGGGTGTAGGGACTGGGAAAGTGGGATTACCTTTCTCACCCATCATA 104856

Qy      1213 AGGGGTCGCAACCAAAAACTCCTATAGTAAAAAGACAGGTTAATAAGAGACAAAACCTAACAAA 1272
             |||||||| |||| ||||||||| |||| ||||| ||| |||| |||||| || |||||| |||
Db    104857 AGGACCACAGCCATTACTCC-ATAACAAAAAGAGAGGTTAAAAGAGAGAAAAGCATAACAAA 104915

Qy      1273 TTTATTTAATCAAAGTTTTTTTACATGACACATGGGAGTCTTCAGAAAATGAAGACCCAAAGACCC 1332
             |||| ||| |||| ||||| |||||||| || ||| ||||||||||||||||| |||| |||||||
Db    104916 TGTATTTAATCACAGTTTTTGTATGTCATGGGAGACCTTCAGAAATGAAGACTGAAGACATTTCACCCATCATA 104975

Qy      1333 AGGGGAAACTGTCTCGTTTTTTTTTTTTGCTTGAGGTTCGATGAAGAATGGCATAGCATGTTAGCCAT 1392
             ||| ||||||| | |||||||||||||||| |||| ||||| |||| ||||| ||| |||| || ||||
Db    104976 AGAGAAAACTGT-GCATTTTTAATGCTTAGGTTAACATTGAAGAATGGGATGACAACCATGTGACAA 105034

Qy      1393 GTAGATTAGACAAAAAGAGGATATAGTTCTAGTTGGTAAAAGGACTCA---GGGGGAAACACAGCA 1449
             ||||||| |||||| ||||| || | ||||| || ||||| |||||||   ||||| |||| ||||
Db    105035 TGTGATTGGGCAAAAAGGAATAATGATGACTAACTAAGAGGAGAGTAAGAGGACTAAGAAACTC 105094

Qy      1450 AGGCCTGTCTATTCAGATTCTTCTTCTTGATCTCTCTCCTAGTAATGCAATTCTTTCTCTC 1509
             |||||| |  || |||||| || | |||||| | | ||||||| ||||| |||||||||||
Db    105095 AGCAAGAGCTGTCCAGATTTTTTCTGGCCTCTC------TGGGTAGCAATTCGTTTCTC 105146

Qy      1510 CTGAGTATGTGGGGGAGGAGGACTCTTCTTCTTCAAGTGAGGGTCTTCATGAAGAAGGAAGTGG 1569
             || | || | ||| ||||||||||| ||||||| |||||| || |||| |||| ||| ||||||
Db    105147 TGTGGCATGGGGCAGGACCCCCTCGGATTGATGTGCTCTTTAAGGGAGAAAGGACAGAGTGC 105206

Qy      1570 CCTTTTAGATTTTATATGGGGTTTGCTTCGGGGAGAGGGGAGAGTTCTAGTTTCTATGACCCATCT 1629
             || ||||| | |||| ||||| |||||||||| || ||| |||||||||| |||||||| || |||
Db    105207 TCTTTCCATCTTTTTATATAGCTTTTGCTTTTGCAGAAGAGGAGTTCTAGTTCGTAGTTTCCATGACCCATCT 105266

Qy      1630 TGGGGAAGAGGAATTCTCGGTTTCTGTGACTTGCTTCATGAAGAAGAGGAGAGTAAGAGGC 1689
             |||| ||||| || ||||||  ||||| || |||||| |||||||||||||| |||||| |
Db    105267 TGAGGAAGAGGAATTCTGATTCAATGATTGCTTTGGAGGAGAAAGGAGAGTACTGTTTCCAAGGC 105326

Qy      1690 AGGAGGGGCAGGAGGATGGGTCAGAGAAAGACTTGGCTGCTTCTCAGGGCTTCGGCTCCCTT 1749
             |||| ||| || || || || |||||||| |||| || ||||| ||||| || ||||| |||
Db    105327 AGGAGGGGCAGGCAAAGCTCAGAAAGA------CGCTTCTGAGGCCTTCCAATCTCTT 105378

Qy      1750 TAGTTCCAAGTACTTCTTTAGCATACCAAAGCACTATACTTTGCATAGTTTTTTCTGAGC 1809
             |||||||||||| |||| || |||| || ||||| ||||| || | ||||||||| ||||
Db    105379 TAGTTCCAAGAA---CTCAGCCATGCCAAAGTGCCATATACTTTTAGGGTACTGTTTCCAAGC 105435

Qy      1810 TCTAACACTGCAATCATGTCTAAACTCTCTCTAT 1841
             |||||||||||| ||| | |||| |||||| |||
Db    105436 CTTAACAAGGGAAACAAAAGTACATCATTTTGTTTAT 105467
```

```
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens genomic DNA
JOURNAL     Published Only in Database (2000)
REFERENCE   2  (bases 1 to 89000)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (22-DEC-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Dec 7, 2001 this sequence version replaced gi:11994966.
FEATURES             Location/Qualifiers
     source          1..89000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="CTD-2005H7"
ORIGIN

Query Match       3.8%;   Score  305.6;  DB 9;   Length  89000;
Best Local Similarity 71.4%;   Pred. No. 2.7e-53;
Matches  494;  Conservative  0;  Mismatches  174;  Indels  24;  Gaps  6;

Qy      1153 CTTCTCTCTCAAAGCTGCAGTAGGGGTGGAAAAACGTGTGATACATTTCTTCCATCATA 1212
             |||||  | |||| |||||    |||||  || || ||||| |||||| |||| |||||
Db     28212 CTTCTGTACCTTGGAGGGGTGTAGGGGACTGGGAAAGTGGGGATACACTTTCTTCACCATGATA 28271

Qy      1213 AGGGGTCGCAACCAAAAACTCCTATAGTAGTAAAAGACAGGTTAATAGAGCAAAACCTAACAA 1272
             |||  | |  |  |   |||| ||  ||  |||||||||||||||  |||||||  |||||||
Db     28272 AGGACCACACGCCATTACTCC-ATAACAAAAGAGAGGTTAAAGAAAAAGCATAACAAA 28330

Qy      1273 TTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAATGAAGACCCCAAAGACCC 1332
             | |||||||||| |||||| | |  ||||||| |  ||||||||||  |||||| |  ||||||||
Db     28331 TGTATTTAATCACAGTTTTATGTGTCATGGGAGCCCTCAGAAATGAAGAGACTCAAAAGACCC 28390

Qy      1333 AGGGGAAAACGTCTCTTTTTTTTTTGCTGAGGTTGCATGAAGACATGTAGGCCAT 1392
             || | ||||| | |    ||||| | || ||||||| ||||||   | |||||||
Db     28391 AGAGAAAACTGT-GCATTTTTATGCTTAGGTTTGATGAAGAAGGCACAAACCATGTAGAAA 28449

Qy      1393 GTAGATTAGACAAAAGGATTATGATTCTAGTGGGTAAAAGGACTCA---GGGGGAAACACAGCA 1449
             |||||||  |||||||||| ||||| | || ||  ||||||| |||   || ||||||||||||
Db     28450 TGTGATTGGGCAAAAGGATATGATTAACAGATAAGACTAAGGAGGGAAACTC 28509

Qy      1450 AGGGCCTGTCTCATTCAGATTCTTCTTCTGATCCTCTCCTATGTATATAGCATTCTTTCCTC 1509
             || |  ||||| |||||||| |||||||||| ||||||         || ||||| |||||
Db     28510 AGCAAGAGCTGTCCAGATTTTCTTCTTCTGGCCTCTC---------TGGGTAGCCATTCGTTTCTC 28561

Qy      1510 CTGAGTATGGGGCAGGACTCTTTCTTCCTGATGAGGGAGGAAGGGAGAGAAGTGG 1569
             | | | ||||| ||||| |  | ||| |||||| | | | ||| |||||||||||
Db     28562 TGTGGCCATGGGAGGGACCCCCTCTTGGATTGAATGGTCTTTAAGGGAGAAGGGGAGAAGTGC 28621

Qy      1570 CCTTTTTAGATTTTTATGGGTTGTTCTTCGGGGAGGAGGGAGGAGAGAGTTTCTTTATGTCTTT 1629
             ||| ||| ||| ||||| |  | |||| |  ||||||  | | || || |||| |    |||||
Db     28622 TCTTTCTTCCATCTTTTTATAGCCTTTTTGCAGAAGAAGTTCTAGTTTCCAGAAGACCATCT 28681

Qy      1630 TGGGGAAGAAGGGAAATCTGGTTTCTGTGACTGGTTTCATGAAGAAAAAGGAGGTAAGAGAGC 1689
             || | ||||||||||||||| |||| |||| | |||| |||||||||| |||| | |||||||
Db     28682 TGAGCAAGAAGGGAAATCTGAGTTTCAATGATTGTTGTTCTTTGAAGAAAAGAGAGTGGGAGGC 28741

Qy      1690 AGGAGGGGCAGGAGATGGCTCAGGAAAAGAGACTTGGCTGTCTGTGAAAAGACTCGAAAGA------ 1749
             ||||||| |||| ||  |||| | |||||||    |  |||||||  |||||| |||  ||
Db     28742 AGGAGGGCAGGCAAAGCTCAGAAAGA------CGCTTCTGAGGCCTTCCAATCTCCTT 28793

Qy      1750 TAGTTCCAAGTACTTCTTAGCATACCAAAAGCACTATACTTTTGGCATATGGGTTTTCTCGAGC 1809
             ||||||| ||       ||||||| ||||||| | ||||||  ||  | ||| |||||| |||||
Db     28794 TAGTTCCCAAGAA---CTCAGCATGCCAAAAGTGCCATACTTAGGGTAGTACTGTTTTCCAAGC 28850

Qy      1810 TCTAACACTCGCAATCATGCTAAAACTCCTCTAT 1841
             ||||||| |  |||  ||||||||||||| ||||
Db     28851 CTTAACAAGGAAGAGAAGAAGTACATTTGTTTAT 28882
```

```
RESULT 13
AL513284/c
LOCUS       AL513284              169972 bp    DNA     linear   PRI 11-JUN-2002
DEFINITION  Human DNA sequence from clone RP11-518D3 on chromosome 1, complete
            sequence.
ACCESSION   AL513284
VERSION     AL513284.12  GI:21425224
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169972)
AUTHORS     White,S.
TITLE       Direct Submission
JOURNAL     Submitted (11-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 13, 2002 this sequence version replaced gi:11873617.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em; EMBL; Sw;
            SWISSPROT; Tr; TREMBL; Wp.; WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-518D3 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..169972
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-518D3"
                     /clone_lib="RPCI-11.2"
ORIGIN

Query Match       3.7%;   Score  299.4;  DB 9;   Length  169972;
Best Local Similarity 69.8%;   Pred. No. 4.8e-52;
Matches  463;  Conservative  0;  Mismatches  191;  Indels  9;  Gaps  4;

Qy      1161 TCAAAGCTGCAGTAGGGGGTGAAAAACGTGTGATACATTTTCTTCCATCATCATAAGGGTGC 1220
             || |  |||| ||||| || |  |||  |||| ||  ||||||||  |||||||| ||| ||
Db    168548 TCTACCCTCGTGTAGGAGGAGGAGGAGAAGGTGTTAATAACTTTTCTTTACCCATCATCAAGGATCAT 168489
```

```
Qy  1221 AACCAAAAACTCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 1280
         ||||||||||||||||||||||| |||||||||||||||||||||| |||||||||||||
Db 168488 GGCCAACACTCCTATAACAAAAAGGCAGAAGAACAGAAAAGCATGACAAAATGTATTTA 168429

Qy  1281 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGGAAGACCCCAAGACCAGGGAAA 1340
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db 168428 ACCAAAAGTTTTATGTGACATGGGAGACCTTGAGAAATAAAGACCCAGGGAAAA 168369

Qy  1341 CTGTCTGTTTTTTTTTTTGCTCGAGGTTCGATCAGAATGATGATCATGTACCATGTAGATTA 1400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 168368 TTGTCTG-TTTTATGCTTAGGTTTGATGAAGAAACTGACACGGCCATATAGAAATGTGATTA 168310

Qy  1401 GACAAAAAGGATATAGATCTAGTGGTAAAGGAACTCA-GGGGGGAAACACAGCAAGGCCTGTCT 1459
         |||||||||||||||||||||||||||||||||||||     ||||||||||||||||||
Db 168309 GACAAAAAGGTATGATCATAATGCTAACACGAGCACGAGAGAAACCCAGCCCAGGCCTGCCT 168250

Qy  1460 ATTCAGATTCTCTTTGATCTCTCTCTTTATGTATAGGCATTCTTTCCTCCTGAGTATGG 1519
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 168249 TT---CGTTCAGATTCTTCGACGCCTCCTATGTAACATTCCTTCCTCCTCGGGGTATGA 168194

Qy  1520 GGCAGGACTCTCTTCAATCAGGGGTCTTCAAGGGGAGAAGGGAAAAGTGGCCTTTTTAGA 1579
         ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db 168193 GGCTGGACTCCTCTGGAAATGAGAATTCTTCAAAGACAGGAGAGAGTGGACTTCCTAGG 168134

Qy  1580 GAGATGGTCAGAAAAGAGACTTGGCCTGCTTCTGAGGCTTCGGCTCTCCTTTAGTTCCAAG 1639
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 168133 GAGTCTGGTTTCTCATTGACTCATTTACTTAGGCCGAGAAAAAGAGACAGGAGATAGGAGG 168074

Qy  1640 GAATTCTTTACATACAAAGACACTATACTTTCATGAAGAAGAAGGAGTAAGAAGGCAGGAGGGCAG 1699
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 168073 GGCTGGACTCCTCTGGAAATGAGAATTCTTCAAGACAGGAGAGAGTGGACTTCCTAGG 168014

Qy  1700 GTAGGAAGAGGTCAGAGAAACCCGTGTTTCTGAGGTCTTCCAATGTCCTTCAGTTCAAAG 1759
         |||||||||||||
Db 168013 GTAGGAAGAGGTCAGAGAAACCCGTGTTTCTGAGGTCTTCCAATGTCCTTCAGTTCAAAG 167954

Qy  1760 TACTTCTTTACCATACAAAAGCACACTATACTTTACTTTTCTGAGCTCTTAAACACTG 1819
         ||||||||||||||||||||||||||||||||||||||||||
Db 167953 TA---CTCAGCATGCCAAGGTGCCATACTTTGGGGTATCAGATTCTGAGCCCCAAACACTC 167897

Qy  1820 CAA  1822
         |||
Db 167896 CTA 167894
```

```
                           30); an attempt was made to resolve all sequencing problems, such
                           as compressions and repeats; all regions were covered by at least
                           one plasmid subclone or more than one M13 subclone; and the
                           assembly was confirmed by restriction digest. The following
                           abbreviations are used to associate primary accession numbers given
                           in the feature table with their source sequence databases: Em., EMBL; Sw.,
                           SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
                           database can be found at
                           http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
                           was generated from part of bacterial clone contigs of human
                           chromosome X, constructed by the Sanger Centre Chromosome X Mapping
                           Group. Further information can be found at
                           http://www.sanger.ac.uk/HGP/ChrX
                           IMPORTANT: This sequence is not the entire insert of clone
                           RP11-435K1 It may be shorter because we sequence overlapping
                           sections only once, except for a short overlap.
                           The true left end of clone RP11-435K1 is at 1 in this sequence. The
                           true left end of clone RP11-342D14 is at 119721 in this sequence.
                           RP11-435K1 is from the library RPCI-11.2 constructed by the group
                           of Pieter de Jong. For further details see
                           http://www.chori.org/bacpac/home.htm
                           VECTOR: pBACe3.6
FEATURES             Location/Qualifiers
     source          1..121720
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="p15.3-15.5"
                     /clone="RP11-435K1"
                     /clone_lib="RPCI-11.2"
     repeat_region   17..90
                     /note="MIR repeat: matches 28. .111 of consensus"
     repeat_region   443. .559
                     /note="MER5A repeat: matches 59. .189 of consensus"
     misc_feature    1986. .2375
                     /note="87 copies 2 mer ta 69% conserved"
     misc_feature    2022. .2428
                     /note="match: GSS: Em:AQ725819"
     misc_feature    5095. .5533
                     /note="match: GSS: Em:AQ442500"
     repeat_region   2544. .2635
                     /note="match: GSS: Em:AQ621701"
     repeat_region   2638. .2932
                     /note="46 copies 2 mer tt 73% conserved"
     repeat_region   3343. .3705
                     /note="AluJo repeat: matches 1. .312 of consensus"
     repeat_region   4022. .4217
                     /note="L2 repeat: matches 2370. .2746 of consensus"
     repeat_region   4624. .4797
                     /note="MIR repeat: matches 8. .208 of consensus"
     repeat_region   5082. .5534
                     /note="MSTB repeat: matches 64. .426 of consensus"
     repeat_region   6039. .6101
                     /note="match: GSS: Em:AQ442489"
     repeat_region   6102. .6395
                     /note="MSTB repeat: matches 1. .64 of consensus"
     repeat_region   6396. .6768
                     /note="AluJo repeat: matches 25. .309 of consensus"
     repeat_region   6979. .7100
                     /note="L2 repeat: matches 2575. .2709 of consensus"
     repeat_region   8009. .8302
                     /note="MER20 repeat: matches 1. .216 of consensus"
     repeat_region   8324. .8823
                     /note="L1ME repeat: matches 943. .1433 of consensus"
     repeat_region   8829. .9032
                     /note="L1ME3 repeat: matches 3815. .4031 of consensus"
     repeat_region   9478. .9746
                     /note="L1ME3 repeat: matches 4058. .4325 of consensus"
     repeat_region   9747. .10084
                     /note="MER44A repeat: matches 1. .332 of consensus"
     repeat_region   10085. .10943
                     /note="L1ME3 repeat: matches 4325. .5188 of consensus"
```

```
repeat_region  10947..11151
               /note="AluJo repeat: matches 86..297 of consensus"
repeat_region  11153..11917
               /note="L1ME3 repeat: matches 5182..6002 of consensus"
repeat_region  11918..12011
               /note="MER81 repeat: matches 2..114 of consensus"
repeat_region  12012..12175
               /note="L1ME3 repeat: matches 6002..6156 of consensus"
repeat_region  14703..15013
               /note="L1ME3 repeat: matches 6002..6156 of consensus"
repeat_region  15013..15834
               /note="AluY repeat: matches 1..311 of consensus"
repeat_region  15834..15923
               /note="L2 repeat: matches 2574..2666 of consensus"
repeat_region  16531..16566
               /note="L2 repeat: matches"
repeat_region  17622..18127
               /note="18 copies 2 mer ca 100% conserved"
repeat_region  18145..18248
               /note="L1MB5 repeat: matches 5679..6174 of consensus"
repeat_region  18392..18695
               /note="FLAM_C repeat: matches 5..108 of consensus"
repeat_region  19551..19659
               /note="AluSg repeat: matches 1..305 of consensus"
repeat_region  19551..19659
               /note="L2 repeat: matches 2549..2662 of consensus"
gene           complement(20811..29961)
               /gene="bA43K1.1"
mRNA           complement(join(<20811..21075,29654..29961))
               /gene="bA43K1.1"
CDS            complement(join(<20811..21075,29654..29886))
               /gene="bA43K1.1"
               /codon_start=1
               /evidence=not_experimental
               /product="bA43K1.1 (novel protein)"
               /protein_id="CAD13520.1"
               /db_xref="GI:17426506"
               /db_xref="SPTREMBL:Q8WX00"
               /translation="MEPQLGPEAAALRPGWIALLLMWSALSCSFSLPASSLSSLVPQV
               RISYNFGRTFLGLDKCNACIGTSICKKFFKEIRSDNWLASHLGLPFDSLLSYPANYS
               DDSKIWARPVEIFRLVSKYQNEISDRRICASASAPKTCSIERVLRKTERFQKWLQAKRL
               TPDLIVQ"
               match: ESTs: Em:AA420823 Em:AA455585 Em:AI796127
               Em:AU121832"
misc_feature   21413..21515
               /gene="bA43K1.1"
               /note="match: GSS: Em:AL013938"
repeat_region  21840..22137
               /note="MIR repeat: matches 42..147 of consensus"
repeat_region  23050..23260
               /note="AluSx repeat: matches 2..299 of consensus"
repeat_region  23316..23609
               /note="MIR repeat: matches 47..262 of consensus"
repeat_region  24752..25050
               /note="AluSx repeat: matches 1..294 of consensus"
repeat_region  complement(25215..25583)
               /note="L2 repeat: matches 2245..2543 of consensus"
misc_feature   complement(25215..25382)
               /gene="bA43K1.1"
               /note="match: GSS: Em:AL013940"
misc_feature   26303..26595
               /gene="bA43K1.1"
               /note="match: GSS: Em:AL013938"
repeat_region  26651..26750
               /note="AluSg repeat: matches 1..293 of consensus"
repeat_region  29031..29118
               /note="L1M4 repeat: matches 5302..5399 of consensus"
repeat_region  31059..31310
               /note="MIR repeat: matches 25..114 of consensus"
repeat_region  31315..31499
               /note="L1M4 repeat: matches 5099..5385 of consensus"
repeat_region  31565..31643
               /note="MER93 repeat: matches 205..397 of consensus"
repeat_region  31649..31998
               /note="MER93 repeat: matches 3..82 of consensus"
repeat_region  ...
               /note="MER57-internal repeat: matches 6808..7145 of
               consensus"
repeat_region  31998..32402
               /note="MER65-internal repeat: matches 3648..4734 of
               consensus"
repeat_region  33415..33449
               /note="MER31-internal repeat: matches 1721..2860 of
               consensus"
repeat_region  33450..33759
               /note="AluSx repeat: matches 1..309 of consensus"
repeat_region  33888..35557
               /note="L1MA2 repeat: matches 4674..6308 of consensus"
repeat_region  35554..37204
               /note="L1M1 repeat: matches 830..2496 of consensus"
repeat_region  37217..37364
               /note="L1M1 repeat: matches -1264..-1116 of consensus"
repeat_region  37382..39552
               /note="L1M1 repeat: matches -1389..828 of consensus"
repeat_region  39638..39764
               /note="L1M1 repeat: matches 418..543 of consensus"
repeat_region  40368..40646
               /note="MER34 repeat: matches 418..543 of consensus"
repeat_region  41671..41824
               /note="MER31-internal repeat: matches 1117..1393 of
               consensus"
repeat_region  41957..42120
               /note="MER93 repeat: matches 248..394 of consensus"
misc_feature   42164..42825
               /note="match: GSS: Em:AQ170009"
repeat_region  ...
               /note="L1M4 repeat: matches 4435..5094 of consensus"
```

```
Query Match   3.7%;  Score 297.6;  DB 9;  Length 121720;
Best Local Similarity 68.6%;           Pred. No.1.2e-51;
Matches 495; Conservative 0; Mismatches 199; Indels 28; Gaps 5;

Qy   1173 TAGGGGTGAAAACGTGGTGATACATTTCTCGTTCCATCATAAGGGTCGCAACCAAAACTCC 1232
          ||| | ||  || ||||| | ||||| |||| || ||| ||||| |||  |||| | |||
Db 102399 TATGGTAGTAGAAAGTGTGTGCCTTTCCTTACCCAGCATTAGGGTCATGGGCGACACCCC  102458

Qy   1233 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAAATTTATTTAATCAAAGTTTTA 1292
          ||| | ||  || ||||||  ||||| |||  ||||||||||||||||| ||||||||||
Db 102459 TATAACAAGGGAAAGGTTAGCAAGAGAAAAGCATAACAAAATTTATTTGATCAAAGTTTTA 102518

Qy   1293 CATGACATGGGAGTCTTCAGAAATGAAACAACCAAAAGACCCAGGGGAAACTGTCTGTTTT 1352
          | |||| |||||| ||||||||||||  || ||||||| |||||   ||||  | |||||
Db 102519 TGTGACACGGGAGCCTTCAGAAATGAAGACCCAAAAGGCCCAGGTCAAA--TGATTATTTT 102576

Qy   1353 TTTGCTGGAGGTTCGATGAAGAAATGGATAGGCATGTGAGCATGTAGAGTTAGAAGGATA  1412
          | ||||  | |||  ||||||||| || | |||||  | || ||   | | ||| | |
Db 102577 TGTGCTTAGGTTTGCATGAAGAAATGGGCAGGTATATGTAGAAAATGGGATAGGACAAAAGGGTG 102636

Qy   1413 TGATCTAGTTGGGTAAAGGACTCAGGGGGAAAACACGCAAGGGCCTGTCTCTATTCAGATTCTTC 1472
          |||||| ||||| |||   || |||| | ||||| | | ||| | ||| |  || ||||| |
Db 102637 TGATCTGTTGGTAATTAAACTGAGGGAGAAAACCCAACAAGTGCTGTGTCTTGTTCAGATTCTCC 102696

Qy   1473 TTGATCTCCTCTCTCTATGTATAGCATTCTTTTCCTCTGAGTATGGGGCAGAGACTCTTC  1532
          || | || ||          ||| |||| ||||| || || ||| |||| || | ||||
Db 102697 TTGGCCATCTC----------TGTGTAGCATTCCTTCCTCTAGGTTATGGGGCAGGATCTTTC 102748

Qy   1533 TTCAATGAGGGTCTTCAGGGAGGAAGGGAGAA--------AGTGGCCCTTTTAGATTTA   1584
          | | |||||||||||||  |||||||||||||        ||||  |||   |||| ||
Db 102749 TGGAATGAGGGTCTTCAAGGGAGGAAGGGAGAAGACTAGAGAGGAGTGATCTTTCCAAGTTTTA 102808

Qy   1585 TGGCTTGCTTGCGGGAGGAGGAGTTCTTAGTTTCTTATGACCCATCTTGGGAGGAGGAAATT 1644
          | |||||||| ||| | ||| |||| || |||| | |||||  | | | ||||| |||||||
Db 102809 TAGCTTGCTTTGCGGAAGTGGAGTCTTACTTTCTCTATGACTCGGCATGAGGGAAGAGGAAATT 102868

Qy   1645 CTGGGTTTCCTGTGACTTGCTTTCATGAAGAAAAGAGGAGTAAGAGGGCAGGAGGGCAGGAGGAT 1704
          |||| ||| | |||||||||||||  ||| ||| ||| |  | ||| |||||||| ||||||||
Db 102869 CTGGTTTCTCATGACTTGCTTT--GGGAGAGAAAGAAGGAAGGATGAGAGAGGAGGGCAGGAGGA 102926

Qy   1705 GGTCAGAAAGAGAACTTGGCGTGCTTCTGAGGGGTTCCGGCTTCCCTTTAGTTCCAAGTACTT 1764
          ||||||| ||||| | ||  | |||||||||| |         |||| || || | ||| ||
Db 102927 GGTCAGAGAGAGAGCAATTACTTCTGAGGA----------CTTCTAATCTGTCAGTTCGGAACTA 102978
```

```
Qy     1765 CTTAGGATACCAAAGCACTATACTTTGGCATATAGGTTTTTCTGAGCTCTAACACTGCAAATC 1824
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102979 CTCACCATACTAAGGTGCGTATTTTGGGGTATTGTTGTTCTAAGCCCAGTGTTATTA 103038

Qy     1825 ATGCTAAACTCCTCTATGACCTTCAAACATTCCACTCGCTGCTTTATTCTTTATTCTTTATGGTTGTGA 1884
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   103039 TGGCTCAAGGATTTTTTCTCTGCAAAAATATGTTTATCCAATTTTATAAATCAGAATGTGA 103098

Qy     1885 TG 1886
            ||
Db   103099 TG 103100
```

RESULT 15

LOCUS   AC010244/c

LOCUS   AC010244   134937 bp   DNA   linear   PRI 15-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-364B20, complete sequence.
ACCESSION   AC010244
VERSION   AC010244.6 GI:15187208
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 134937)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 134937)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 134937)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 134937)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   On Aug 15, 2001 this sequence version replaced gi:12545284.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.6% of Sequence;
            Estimated Total Number of Errors is 0.4.
            STS Content:
            SHGC-79385 G52178.

FEATURES   Location/Qualifiers
     source   1..134937
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-364B20"

ORIGIN

Query Match   3.7%; Score 293.8; DB 9; Length 134937;
Best Local Similarity   72.8%; Pred. No. 7.2e-51;
Matches 481; Conservative 0; Mismatches 157; Indels 23; Gaps 7;

```
Qy     1162 CAAAGCTGCAGTAGGGCGGTGAAAAACGTGTGATACATTTTCTTCCATCATAGAGGGTCGCA 1221
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    22463 CTATCCAGCTGTACAGGTAGTAGTGTGATATGTTTTTTCTCACCTATAATAGAGTCACA 22404

Qy     1222 ACCAAAAACTCCTATAGTAAAAGACAGGTTAATAAGACGAAAAACCTAACAAAATTTATTTA 1281
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    22403 GCTGACAATTCCTATAACAAAAGAGATAAGGTTAATAAGAGGAAAAAGCATACACACTTTTATTTA 22344
```

```
Qy     1282 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAAATGAAAGACCCAAAGACCAGGGGGAAAC 1341
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    22343 TCAAAGTGTTACTTGATACAAGAGCCTTCAGAAAATGAAGACCCAAGACCAGAGAAAAC 22284

Qy     1342 TGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGGCACGTAGGATTAG 1401
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    22283 TG--CCCTTTTTTATGCTTATGTTTAGTTGAAGAAGAATGGACAGCGCCAGGTGGAAAATGTGATTGG 22225

Qy     1402 ACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAACACAGCAAGCCTGTCTA 1460
            |||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    22224 ACAAAAGGATATAATTAAAGGTGAGGAAGAATGAGGGGGAAAAAACCCAGCAAGGCCTGCCTG 22165

Qy     1461 TTCAGATTCTTCTTGATCTCTCTCTCTGTCTATAGCATTCTTCCTCCTCCTGAGTATGGG 1520
            ||||||||||||||||||||||||||||||||||||||||||||||
Db    22164 TTCAGATTCTTCTTCTTGACTTCTTTCTGT-------AACATTCCTTCTTTCTGGGGTTATAGG 22113

Qy     1521 GCAGGACTCTTCTTCAAGAGGTCTTCAAGGAGAAGGAGAAAGTGGCCTTTGGCCTTTGAGAT 1580
            |||||||||||||||||||||||||||||||||||||||||||||      ||||
Db    22112 GCAGGATTACTTCTGGAATGAGGGTGCTTGGAAATGGGGGAGAAGGAGAA---GCGACTTTTTGGC 22056

Qy     1581 TTTATTGGCTTGCTTCGGGGGGAAGGAGGAGTTCTAGTTCTTTCTTATGACCCATCTTGGGGGAGGAGG 1640
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    22055 TTTATGGCTGACTTCGGGGAGCGGGGGTTCTAATTTCTATGAAACTGCCTTGCCAGGAGAGT 21996

Qy     1641 AATTCTGGTTTCTGTGACTTGCTT--TCATGAAGAAGAGAGGAGTAAGAGGCCAGGAGGGCAG 1699
            |||||||||||||||||| |||||  ||||||||||||||||||||||||||||||||||
Db    21995 AATTTTGGTTTTGGTTGAATTGCTTCAGGGGGAGAAAGATGGTTGGGAGGAAGAGAACAG 21936

Qy     1700 GAGATGGGTCAGAAGAGACTTGGCTGCTTTCCAGGGTTCCGCTCTCCCTTTAGTTCCAAG 1759
            ||||||||||||||||||||||| ||||||||||||||||
Db    21935 GAGAAGATCAGAACAGA-----CTTTGCTTTTGAGGCCTTCCAATCTCCTTAGGT---CA 21885

Qy     1760 TACTTCTTTAGCATACCAAAGCACTATACTTTGGCATATAGGTTTTTCTGAGCCTTAACACTG 1819
            |||||||||||||||||||||| ||||||||||||||||||||||||||
Db    21884 AAGTAAATCAGCATACCAAAGTAATGTCTTCGTATTCTGAGCCCAAACACAG 21825

Qy     1820 C 1820
            |
Db    21824 C 21824
```

Search completed: September 11, 2004, 16:46:42
Job time : 19875 secs